MATHEMATICA Policy Research

REPORT

FINAL REPORT

Baseline Report for the Impact Evaluation of the SI-EITP model

February 28, 2019 (revised April, 2019)

Ivonne Padilla Espinosa
Larissa Campuzano
Camila Fernández

Submitted to:
Millennium Challenge Corporation
875 Fifteenth Street NW
Washington, DC 20005-2221
Project Officer: Anne Pizer
Contract Number: MCC-13-BPA-0040/MCC-14-CL-0003

Submitted by:
Mathematica Policy Research
P.O. Box 2393
Princeton, NJ 08543-2393
Telephone: (609) 799-3535
Facsimile: (609) 799-0005
Project Director: Larissa Campuzano
Reference Number: 40349.ER4

This page has been left blank for double-sided copying.

CONTENTS

TABLES

FIGURES

This page has been left blank for double-sided copying.

LIST OF ACRONYMS

CDE	Consejo Directivo Escolar
CLS	Classroom Life Scale
DDEs	Dirección Departamental Educativa (Department's Education Offices)
FOMILENIO II	Millennium Challenge Fund of El Salvador
ICT	Information and Communications Technology
INSAFORP	Salvadoran Institute of Vocational Training (Instituto Salvadoreño de Formación Profesional)
MCC	Millennium Challenge Corporation
MINED	Salvadoran Ministry of Education
NIE	Número de Identificación del Estudiante
PAES	Prueba de Aprendizaje y Aptitudes para Egresados de Educación Media
PNFM	Programa Nacional de Formación de Maestros
SAE	Sistema de Administración Escolar (School Information Management) System
SIRAI	Sistema Institutional de Registro Académico e Institucional (Academic Registry System for Secondary Education)
TEAs	Technical education assistants

This page has been left blank for double-sided copying.

I. INTRODUCTION

The Integrated Systems of Full-Time Inclusive Schools (SI-EITP for its acronym in Spanish) intervention aims to expand and improve the learning opportunities offered to children in all primary and secondary grades in El Salvador in each integrated system's geographic territory. The SI-EITP model is based on a school-level model (full-time inclusive schools) that the Ministry of Education (MINED) in El Salvador started implementing in selected schools in 2015 and that it later modified for implementation across a group of neighboring schools (Integrated Systems). In this study, we evaluate the SI-EITP model that the Millennium Challenge Fund of El Salvador (FOMILENIO II for its acronym in Spanish) is implementing in the eight departments of the Coastal Region in El Salvador with funding from the Millennium Challenge Corporation (MCC).

The implementation of the SI-EITP model in the Coastal Zone is one of several interventions supported by the MCC under the Second Compact signed with the Government of El Salvador in 2014. The SI-EITP intervention is one of the two sub-Activities implemented under the Education Quality Activity that aims to improve the quality of El Salvador's national education system, with funding of nearly $85 million invested in both sub-Activities. The other sub-Activity is the Strengthening of the National Education System, which is designed to improve the effectiveness and quality of El Salvador's national education system. MCC hired Mathematica Policy Research to evaluate the two sub-Activities under the Education Quality Activity. The evaluation of the SI-EITP intervention includes both impact and performance studies.

In this report, we present baseline findings for the impact evaluation of SI-EITP. The impact evaluation is based on random assignment of eligible systems into two groups: an intervention group that is implementing the SI-EITP model with FOMILENIO II's funds and a control group that is not implementing the SI-EITP model. Data for the report come from principal, teacher, and student surveys and structured classroom observations collected in October 2017. FOMILENIO II's implementation in the 45 Integrated Systems started in 2018. In particular, analyses presented in this baseline report aim to:

1. Describe the context in which the SI-EITP model is undergoing implementation, including the characteristics of the study schools, teachers, and students in 2017, in the school year before FOMILENIO II's implementation of SI-EITP began
2. Assess the degree to which random assignment produced equivalent groups based on measurable characteristics for schools, teachers, and students in 2017

The main finding is that both groups had similar characteristics for schools, teachers, and students. Although we found a few significant differences, there is no evidence that one group had better characteristics than the other in general.

In this chapter, we present FOMILENIO II's implementation plans for SI-EITP. Then, we discuss the evaluation questions, design, and data sources. In Chapter II, we present results related to the equivalence of our treatment and control groups on background characteristics of schools, teachers, principals, and students. In Chapter III, we discuss equivalence on teacher's use of class time and learning activities in the base year. We then outline in Chapter IV the

baseline equivalence results on students' use of time in and out of school, along with perceptions of teaching practices and of classroom climate and safety in the community. In Chapter V, we describe the baseline equivalence results on education outcomes. Finally, in Chapter VI, we summarize the results and conclude that the random assignment process produced comparable groups.

A. FOMILENIO II's implementation of the SI-EITP intervention

The SI-EITP model organizes neighboring schools of all grade levels into an integrated system (or cluster) in which representatives from the schools in the system work together to develop joint action plans, optimize and share resources, exchange expertise, and foster the involvement of families and the community. The goal of the model is to expand and improve the learning opportunities offered to children in all primary and secondary grades in each system's geographic territory and, ultimately, to improve students' labor market outcomes. Even though the SI-EITP model targets students in all grades, FOMILENIO II's implementation of the SI-EITP model focuses mainly on grades 7 to 12. The main objectives of implementation are to (1) increase enrollment and continuation in and completion of third-cycle (grades 7 to 9) and secondary (grades 10 to 12) education; (2) extend the length of the school day to increase the total number of hours per week from 25 to at least 36 hours (and up to 40) per week in grades 7 to 9; (3) enhance students' development through the use of active-learning methods in core subjects and the use of information technology; and (4) improve school management and the sharing of resources across schools in the Integrated Systems.[1]

MINED has been implementing the SI-EITP model in selected schools since 2015. With MCC funding, FOMILENIO II is implementing the SI-EITP model in eight departments covering 33 municipalities of the Coastal Region. FOMILENIO II's implementation of the SI-EITP model differs somewhat from previous implementations. For example, FOMILENIO II identified the infrastructure needs of schools that will be addressed with implementation of the SI-EITP model. In addition, implementation includes the development of reading communities for all grade levels and the strengthening of English-language teaching for third-cycle and secondary schools. FOMILENIO II has estimated that funds are available to implement the SI-EITP model in 45 of 147 potential integrated systems identified by MINED in the Coastal Region (Figure I.1). FOMILENIO II's resources are mainly focused on education in third-cycle (grades 7 to 9) and secondary schools (grades 10 to12). In El Salvador, secondary schools offer two types of track: a general track in grades 10 and 11 that allows students to complete a general baccalaureate and a technical track in grades 10 to 12 that allows students to complete a baccalaureate with a technical specialization. Each Integrated System will designate one school as *centro educativo integral* (CEI)— a school with enhanced infrastructure and resources that will be accessible to all the other schools in the Integrated System.

[1] The Evaluation Design Report for the Education Quality Activity describes the SI-EITP model and its previous implementations in more detail (Campuzano et al. 2018).

Figure I.1. Geographic distribution of the Integrated Systems participating in implementation of the SI-EITP model of FOMILENIO II, by department

FOMILENIO II's implementation of the SI-EITP model differs from MINED's implementation in the following ways:

- FOMILENIO II first conducted characterization studies in schools selected to implement the SI-EITP model in order to assess the needs of each system;

In addition, the intervention includes the following components:

- Redesign the English-language curricula and provide training to strengthen English-language teaching for third-cycle and secondary schools;
- Establish reading communities for all grade levels;
- Offer new technical programs and certificate programs[2] in secondary schools;
- Construct or rehabilitate school facilities.

In 2018, FOMILENIO II contracted with the organizations that were selected to implement the components of the SI-EITP model. As FOMILENIO II's plans for implementation developed, the distinction between the two sub-activities under the Education Quality Activity—SI-EITP sub-Activity, and the Strengthening the National Education System sub-Activity—became less clear. Therefore, even though the 11 components discussed below describe the implementation of the Education Quality Activity, some of the components are implemented at the national level and others only in the 45 Integrated Systems that are part of the intervention.

[2] Certificate programs (diplomados in Spanish) offer courses to general baccalaureate students focused on skills that can improve their employability such as graphic design, English language, and communications. These certifications do not require the additional year that the technical programs require.

The Evaluation Design Report (Campuzano et al. 2018) presents a detailed description of the components. Below, we present a high-level summary of the components under implementation of the Education Quality Activity.

Component 1: Professional development for specialists and teachers. The main goal of Component 1 is to strengthen the content and pedagogical knowledge and technological and social competencies of teachers in schools within the 45 Integrated Systems in which the model is being implemented. The main activities follow:

- Provide training programs for specialists (teacher trainers) in the SI-EITP model, integrative methodologies, and evaluation of learning
- Provide training for teachers of children in grades 7 to 12 focused on content knowledge, such as in mathematics, language, biology, chemistry, physics, and social science. The training will follow the national teacher professional development program (*Programa Nacional de Formación de Maestros*, or PNFM), but specialists trained through FOMILENIO II will deliver the training and will teach integrative methodologies and evaluation of learning.
- Provide training in language, mathematics, and strategies to teachers of multigrade classrooms of children in grades 1 to 6.
- Provide training in socioemotional development skills, technological literacy, and gender inclusion to teachers of children in grades 1 through 12
- Design and implement workshops and technical assistance focused on improving principals' school management skills
- Redesign the training modules of the six content areas of the PNFM as needed at the national level; all the other activities included in Component 1 are being implemented in the 45 Integrated Systems.

Component 2: Strengthening pedagogical technical assistance and school management. The main goal of Component 2 is to strengthen and provide timely and effective technical assistance on pedagogy and school management in the 45 Integrated Systems. The main activities are follow:

- Hire 30 technical education assistants (TEAs) and train them to provide technical and pedagogical assistance to the 45 Integrated Systems selected for implementation.
- Offer training to strengthen the management skills of MINED's central office staff and the staff of the 14 Department's Education Offices (DDEs for its acronym in Spanish).
- Design and implement a software to track technical assistance visits and link it to the Salvadoran Management Educational Information System (SIGES for its initials in Spanish). SIGES is the computer system that will automate processes, improve the quality of the data collected by MINED, and facilitate the monitoring of educational indicators

Component 3: Governance and participation of SI-EITP. The main goal of Component 3 is to improve the governance of the 45 Integrated Systems. The main activities follow:

- Design and produce materials on management of SI-EITP, the technical and pedagogical teams and the five school committees: Consejo Directivo Escolar CDE, Consejos de Dirección, Consejos Docentes, Consejo de Estudiantes, Consejo de Familia.
- Train TEAs and staff from the 14 DDEs in the management of SI-EITP and the establishment of the five school committees.
- Support the establishment of the 45 Integrated Systems. For the 45 Integrated Systems, offer school management, leadership, and community participation workshops to (1) school management councils, (2) principal councils, (3) teacher councils, (4) technical and pedagogical teams, (5) student councils, and (6) parent councils.

Component 4: Strengthening MINED's technical capabilities in terms of learning and curricular evaluation. The main goal of Component 4 is to strengthen the technical capabilities of MINED in curricular and learning evaluation and to facilitate student participation in international standardized tests. The main activities follow:

- Offer training to specialists and MINED staff in areas associated with the application of standardized tests, such as evaluation models, preparation of items (questions for the test), development of tests, and communication of results
- Provide MINED with bibliographic resources and specialized software to facilitate the processing of data from standardized tests
- Support the application of the Latin American Laboratory for Assessment of the Quality of Education (LLECE for its acronym in Spanish) test in 2019
- Redesign the curricula and materials for language and literature with an emphasis on a competency-based approach
- Offer technical assistance to the TEAs in the 45 Integrated Systems to strengthen their technical capabilities in the use of standardized test results

Component 5: Development of curricula and implementation of the new technical offerings in technical programs and certificate programs (*diplomados*) in general programs. The main goal of Component 5 is to improve alignment between the productive sector's needs and the technical programs offered in secondary schools in the Integrated Systems. The main element of this component is the implementation of new offerings in selected general and technical programs within the Integrated Systems. The main activities follow:

- Complete the curricular designs of the new technical secondary options (services, agriculture, and industry and innovation)
- Design the teacher materials for implementation of the new options for technical secondary education; design teacher training materials, along with the new certificates for the general programs
- Design and deliver the technical education assistance program for the principals and teachers charged with implementing the new technical options
- Establish a curricular evaluation strategy for the three new technical offerings

- Design a sustainable model for new technical offerings for MINED's introduction at the national level

Component 6: Extension of the school day and of life and work skills training for students in third-cycle schools. The main goal of Component 6 is to extend the school day to total at least an additional 11 hours per week in grades 7 to 9 in the Integrated Systems. The main activities follow:

- Design workshops that will allow third-cycle schools (grades 7 to 9) to extend the school day hours with at least three hours per week of the extended school schedule focusing on English, two hours on sports, and six hours on workshops to teach work and life skills
- Hire and train staff, organize the school schedule to include the workshops, develop the workshop materials, and offer the workshops

Component 7: Strengthening English-language teaching for third-cycle and secondary schools. The main goal of Component 7 is to improve oral and written comprehension of the English language among students in the Integrated Systems. The main activities follow:

- Redesign the English-language curriculum for third-cycle (grades 7 to 9) and secondary (grades 10 to 11) schools.
- Provide training to English-language teachers in the 45 Integrated Systems. During the first year of implementation, teachers will take English-language classes (tailored to their level). In the second year, teachers train in practices for teaching English.
- Provide tablets to the schools in the 45 Integrated Systems for use in English-language instruction in grades 7 to 9.
- Design and provide English-language materials to the schools in the 45 Integrated Systems.

Component 8: Development of reading communities in SI-EITP. The main goal of Component 8 is to promote reading and to develop communication skills among students in the 45 Integrated Systems. The main activities follow:

- Establish reading communities with students in all grades (grades 1–11), focusing on the primary grades and progressively expanding to the secondary grades.
- Establish two libraries in each of the 45 Integrated Systems. Form 90 library committees and conduct training in library use.
- Train teachers, student leaders, committees, and families in promoting reading. Form groups of student leaders to promote reading.
- Deliver tablets to third-cycle schools for reading.

Component 9: Implementation of gender equity and equality policy. The main goal of Component 9 is to strengthen the implementation of inclusive, nonsexist teaching practices. It also seeks to improve MINED's response in cases of gender violence and to promote measures

that eliminate gender-based inequalities and discrimination in school. Component 9 is to be implemented at the national level. The main activities follow:

- Train 300 teachers to become specialists and 1,400 teachers in nonsexist teaching practices
- Train MINED staff from its gender unit as well as key personnel in nonsexist education

Component 10: Information systems governance and data quality assurance. The main goal of Component 10 is to establish and operate an information system that automates processes, improves the quality of data collected by MINED, and facilitates the monitoring of educational indicators. The main activities follow:

- Design and implement a functional web application based on the National Education System, which integrates the information from all divisions of MINED. This new application is referred to as SIGES. The core modules of the system will be information collected from schools, teachers, and students. Complementary modules will include financial information from schools as well as other educational statistics.
- Offer assistance to MINES in the development of technical capacity for the maintenance and effective use of SIGES and its data.

Component 11: School infrastructure. The main goal of Component 11 is to offer more options for secondary education through appropriate equipment and infrastructure in the 45 Integrated Systems. The main activities follow:

- Construct or rehabilitate school facilities and provide equipment to offer an appropriate physical learning environment to the 45 Integrated Systems. In each system, one school (*centro educativo integral*) will be prioritized and will serve as the resource center for the other schools in the system.

B. Evaluation questions and design

In this section, we summarize the research questions and the design of the SI-EITP impact evaluation (a detailed discussion appears in Campuzano et al. 2018). The goal of the impact evaluation is to estimate the effect of the SI-EITP model on key outcomes. The main research questions follow:

1. What is the impact of the SI-EITP on dropout, continuation to the next grade, and graduation from secondary education?
2. What is the impact of the SI-EITP on the academic performance of students as measured by the Prueba de Aprendizaje y Aptitudes para Egresados de Educación Media (PAES) in grade 11?[3]

[3] If MINED implements national testing on another relevant grade level, we will consider assessing the impact on that test.

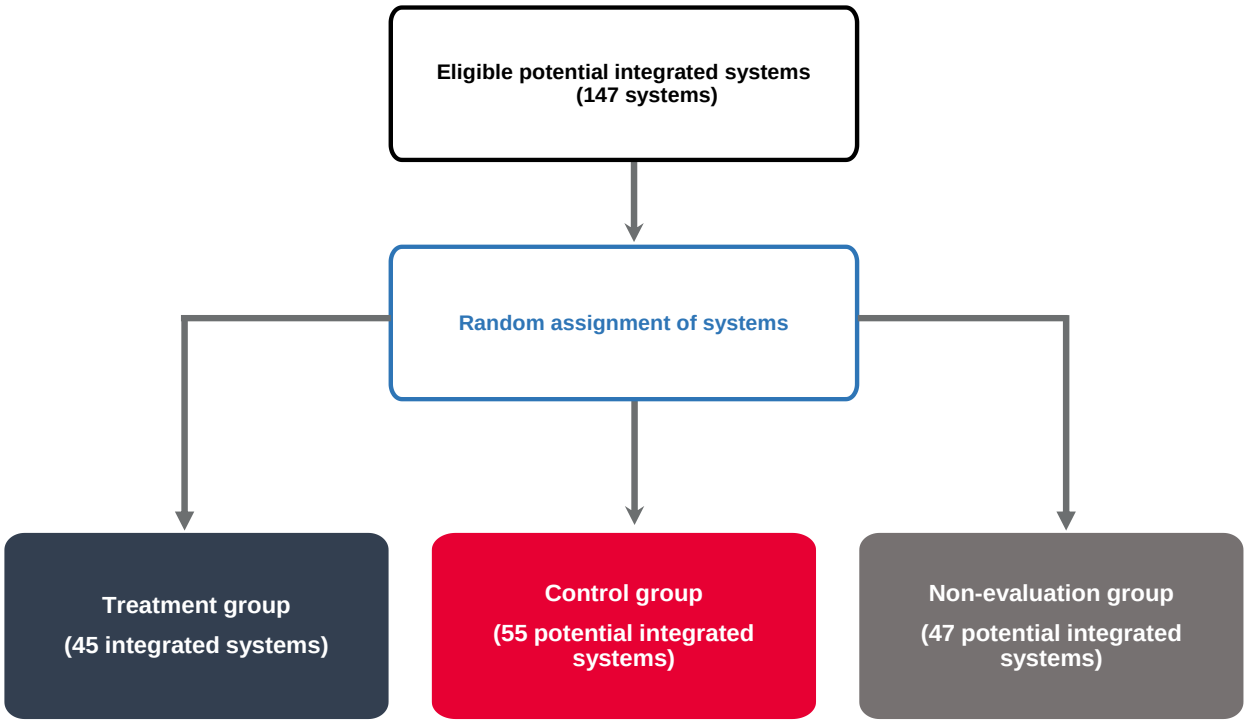
3. What is the impact of the SI-EITP on the length of the school day? What is the impact of the SI-EITP on the quality of education as measured by time on-task?
4. What is the impact of the SI-EITP on postsecondary education enrollment, employment, and income?[4]

The impact evaluation consists of a randomized controlled trial, the most rigorous evaluative framework for assessing program impacts. At the end of 2014, we randomly assigned the eligible systems (147 systems) in the Coastal Zone of El Salvador to one of three groups. The treatment group (45 systems) is being offered all of the components of the SI-EITP model as implemented by FOMILENIO II, whereas the control (55 systems) and nonevaluation (47 systems) groups are only receiving the services already provided by MINED in each region (Figure I.2). The evaluation, however, includes only systems in the treatment and control groups; due to cost constraints, we are not collecting data on the schools in the systems of the nonevaluation group. Yet, all groups are receiving the components or activities of FOMILENIO II's Education Quality Activity that are being implemented at the national level. To allocate eligible systems to the treatment and control groups, we considered two factors: (1) the administrative department in which the systems are located and (2) the presence of at least one secondary school in each system. The use of the administrative department for random assignment of systems guaranteed that we assigned systems to either the treatment or control group in all eight departments.

The randomized design relies on comparing outcomes between the treatment and control groups to estimate the impacts of the SI-EITP model that is implemented in the treatment group but not in the control group. Therefore, this impact evaluation of the SI-EITP model will assess only the joint effect of the components implemented in schools in the treatment group, but not the effect in control group schools. In other words, we will not be able to separate the effects of each component implemented as part of SI-EITP model. Furthermore, given that some components are implemented at the national level and affect both the treatment and control groups, we will not be able to evaluate the effect of the components implemented at the national level, such as the implementation of the new information system for educational outcomes.

[4] This question is contingent on MCC's exercise of the option to collect postsecondary data.

Figure I.2. Random assignment of potential systems



C. Data sources and indicators

The design report details all data sources for the impact evaluation of the SI-EITP model (Campuzano et al. 2018). This baseline report relies on two main data sources: (1) in-school surveys of principals, teachers, and students as well as structured classroom observations collected by a third party and (2) MINED's student-level administrative data from third-cycle and secondary schools, along with test scores from the PAES, which is a national achievement test administered in grade11. Below, we provide a brief description of the data sources.

1. In-school data collection

Primary data collected in schools came from principal, teacher, and student surveys, along with structured classroom observations in 190 schools. In-school data collection took place in October 2017, before FOMILENIO II began implementation of the SI-EITP model. El Salvador's National Statistical Office (DIGESTYC for its acronym in Spanish) conducted data collection field work, with close support and oversight from Mathematica. Follow-up data collection will take place in the same schools that made up the baseline sample as well as in the new secondary schools established in the Integrated Systems.

Principal survey. We interviewed the principal in each school in the sample. Principals from study schools provided information about the available facilities and infrastructure in schools. We also asked the principals about schools' organization, for example, the number of hours that the school offered activities (including instructional time and extracurricular activities), the activities offered in the extended time period, and the resources shared with nearby schools.

Teacher survey. We interviewed teachers to learn about their background, participation in professional development activities, and self-reported teaching practices in the classroom. Teachers reported on their participation in the preparing the pedagogical proposal[5] and the extent to which resources are shared across schools. Finally, the survey of teachers in third-cycle schools asked for information about the teachers' participation in activities offered during the extended school day and about the alignment of the course they are teaching and their area of study.

Student survey. We interviewed students to understand students' time use in academic and nonacademic activities and to gather information on their participation in extension activities offered throughout the Integrated Systems. Students reported on time spent on academic/ instructional activities as well as on time spent on sports and recreation or vocational activities with or without teacher supervision, including recess time. Students also reported on their school engagement, perceived support from teachers, gender biases in the classroom, and school safety.

We used the above data to construct measures of students' perceptions of teacher academic support and behavioral and emotional engagement in school. We used items from the Classroom Life Scale (CLS) questionnaire and the methodology developed by Van Ryzin, Gravely, and Roseth 2009 to construct the following subscales:

a. **Teacher academic support subscale.** To construct the classroom life subscale of teacher academic support, we used the students' answers to four questions about academic support, such as *my teachers want me to do my best in schoolwork*, measured with a 5-point Likert-type scale with 1 as never and 5 as always.

b. **Behavioral and emotional engagement subscales.** To construct the behavioral engagement subscale and emotional behavioral engagement subscale, we used the answers to 20 questions about effort and attention (behavioral engagement) and interest and enjoyment (emotional engagement), measured with a 5-point Likert-type scale with 1 as never and 5 as always.

c. **Classroom climate and teaching practices.** Relying on a 4-point scale from completely disagree to completely agree, we used the answers to 10 items about how students felt in the classroom. We constructed binary variables for each item if the student agreed with statements such as *I feel comfortable sharing my own perspectives and experiences in class.*

d. **Students' perception of teaching practices.** Relying on a 4-point scale with 1 as a few teachers and 4 as all teachers, we used the answers to 10 items about the proportion of teachers who performed teaching practices that aimed to encourage or engage students. We constructed binary indicators for each item to show the percentage of students reporting that most of their teachers used these teaching practices.

[5] The pedagogical proposal is a document prepared by the school that adapts the content of the national curriculum taking into account the learning needs of students and the historical and territorial context in which they are living and describes the school work plan.

e. **Perception of prioritized teaching practices.** We used the answers to 10 items about how important it was for teachers that their students performed activities such as understand the readings, study and do homework, memorize information.

f. **Perception of community safety.** We used the answers to three items about students' perceptions of safety at school, going to school, and in their community. We constructed a binary indicator for each item to show the percentage of students who felt safe.

In Table I.1 we summarize the contents for the three surveys conducted at baseline. Detailed information on students' constructs appears in Appendix A, Table A.1.

Table I.1. Description of the surveys administered in school in base year

Domain	Survey contents
Principal survey	
Demographics and background information	Gender, age, education level, work experience, etc.
School shifts and length of school day	Number of shifts the school offered, length of the school day, participation in the Integrated Systems
Pedagogical decision making	Participation in pedagogical proposal
Parent and community school involvement	Percentage of parents of students in classrooms who were involved in school activities
Professional development	Principal's participation in professional development activities in 2017 or previous years
Resources, infrastructure, and safety	Available facilities and infrastructure in schools
Teacher survey	
Demographics and background information	Gender, age, education level, work experience, etc.
Self-reported teaching practices	Teacher's reported frequency in performing some general teaching practices such as presenting information on the blackboard, solving exercises, and promoting group discussion
Parent and community school involvement	Percentage of parents of students in teacher's classrooms who were involved in school activities
Pedagogical decision making	Participation in pedagogical proposal
Professional development	Teacher's participation in professional development activities in 2017 or previous years; teacher reporting if he/she received pedagogical support or coaching
Student survey	
Demographics and background information	Gender, age, grade attended
Students' use of time on weekdays	Students' reported time spent on academic/instructional activities as well as time spent on sports and recreation or vocational activities with or without teacher supervision, including recess time
Students' perception of teacher support and engagement in schools	Students' perceived academic support from teachers, self-reported involvement in academic activities, reported frequency in performing learning activities such as science experiments, asking questions of teachers, etc.
Students' perception of teaching practices and gender equity	Students' reported percentage of teachers who performed teaching practices aimed at encouraging or engaging students in the classroom and students' perception of gender equity

TABLE I.1 *(Continued)*

Domain	Survey contents
Students' perception of school and community safety	Students reported perception of safety in the community and the frequency of problems such as theft of belongings and fights occurring in school
Expectations about academic achievement and interests	Students' expectations about completing secondary school and pursuing a technical or professional career; students' interest in the workshops

Structured classroom observation. In addition to administering the surveys, we conducted a structured classroom observation by using the Stallings Classroom Snapshot instrument. The Stallings observation consists of 10 separate observations or snapshots[6] made at regular intervals throughout a class period. In each snapshot, the observer scans the room in a 360-degree circle starting with the teacher and codes in detail the following key aspects of classroom dynamics: the teacher's use of class time, the instructional activities taking place, the materials used in the classroom, and the teacher's interaction with students. With this instrument, we created the following four measures:

1. Share of class time during which teachers engaged in the following core instructional activities: reading aloud, demonstration/lecture, discussion/question and answer, practice and drill, monitoring copying, and monitoring seatwork
2. Share of class time during which teachers engaged in three broad categories of activity: instruction, classroom management, or other activities not related to teaching (off-task)
3. Share of class time during which teachers used the following learning materials: blackboard, books, notebooks, didactic material, information and communications technology (ICT), or laboratory equipment
4. Share of class time during which students were engaged as the teacher conducted an instructional activity and share of class time during which students were engaged in off-tasks behaviors

Evidence suggests that the way teachers use time and employ academic activities in classrooms is associated with student achievement outcomes. For example, Stallings and Knight (2003) applied the Stallings instrument in several U.S. school districts and observed high-performing schools over several decades. They found that teachers whose students showed gains on various standardized achievement tests devote an average 85 percent of class time to instruction. Moreover, these same teachers spend 50 percent of total class time on "active" instructional activities (where teachers engage with students through lecture and explanation and question and answer interaction) and not more than 35 percent of total class time on "passive" instruction (such as monitoring copying or seatwork). These teachers also dedicate 15 percent or less of class time to organizing and managing activities such as distributing papers, taking attendance, and explaining the week's schedule. In addition, these teachers were never observed off-task; that is they were never observed out of the classroom, socializing with other adults or visibly not engaged with the activities performed by the students in class. Drawing on the correlation found between teachers' use of time and students' gains observed in achievement

[6] The classroom snapshot records teachers, students, and activities performed in a classroom as if they were being photographed during one brief instant, hence the word "snapshot." Each "snapshot" observation lasts for 15 seconds.

tests, Stallings and Knight suggest that the above distribution of class time among instruction, management, and off-task activities is a benchmark for "good practice." Stallings and Knight also observed that these effective teachers do a good job of keeping students engaged in learning throughout the class.

The Stallings observation tool offers several advantages. For instance, it generates quantitative measures on the interaction of teachers and students in the classroom across different grades and subjects with a high degree of inter-rater reliability among observers with relatively limited training. Furthermore, the instrument is language- and curriculum-neutral, so that results are comparable across different types of schools and country contexts. However, one limitation is that the instrument does not capture content knowledge or adherence to a given curriculum (Schuh Moore, DeStefano, and Adelman, 2010).

2. Student-level administrative data from MINED

To construct the student outcome indicators such as enrollment, dropout, progression in school, and academic achievement measured by PAES, we relied on student-level data from MINED's administrative systems for all the schools in the study sample. For the baseline analysis, we used student data for school years 2016 to 2018.

Student-level educational indicators. We used student-level data to construct measures of student enrollment, dropout, repetition, and grade progression for third-cycle and secondary schools. We present the definition of these indicators in Table A.1. MINED's student-level data are available in two information management systems that use a unique student identification number (Número de Identificación del Estudiante or NIE for its acronym in Spanish) provided by MINED. The first information management system is called the School Management System (Sistema de Administración Escolar or SAE for its acronym in Spanish), and it houses data for students in third-cycle grades. The SAE is mainly an initial enrollment registry and does not collect information on students' status at the end of the school year. We used two consecutive years (school years 2017 and 2018) of student data for all the grades in the study schools to match students by NIE and to identify the students who were repeating grades or had progressed to the next grade. Similarly, we constructed interyear dropout (within school) by identifying students who were enrolled in one school year but not enrolled in the next school year (neither repeating nor progressing). The outcome indicators that we cannot construct with the SAE pertain to whether students were passing or failing a grade because the system does not provide such information.

The second information management system is called the Academic Registry System (SIRAI for its acronym in Spanish), administered by the Office of Accreditation at MINED. SIRAI houses enrollment data for students in secondary grades as well as information on the status of students at the end of the school year. With these data, we constructed student-level indicators on dropout, passing a grade, repeating a grade, and graduating from secondary school (within school). In addition, we constructed the indicator for the transition from grade 9 to 10 by using two consecutive years of SAE (grade 9) and SIRAI (grade 10) data. The transition indicator identified whether grade 9 students made the transition to grade 10.

Student test scores. Each year, all students enrolled in grade 11 in El Salvador take the Prueba de Aprendizaje y Aptitudes para Egresados de Educación Media (PAES) test, a national

test of language, mathematics, science, and social studies. Using student-level scores from PAES for all students in the sample schools, we measured academic achievement in base year 2017.

3. Study sample

Our evaluation sample consists of 100 Integrated Systems, 55 in the treatment group and 45 in the control group (Figure I.2). To draw the sample for the in-school baseline data collection, we randomly selected one third-cycle school and one secondary school from each Integrated System in the evaluation sample. However, one Integrated System in the control group did not include third-cycle schools, and 26 Integrated Systems did not include secondary schools (21 in the treatment group and 5 in the control group).[7] For this baseline study, we collected in-school data in 116 third-cycle schools (54 schools in the treatment group and 62 schools in the control group) and 74 secondary schools (34 schools in the treatment group and 40 schools in the control group).

We developed the following protocols to select the classrooms to be observed and the principals, teachers, and students to be surveyed:

- We interviewed the principal in each school in which we collected data.

- We developed a protocol to select the classes in which the Stallings observations would take place. We directed enumerators to choose one class per subject area (language, mathematics, and science) in grades 7 to 9 (third-cycle schools) and grades 10 and 11 (secondary schools). We randomly selected the shift and order in which classes would be observed. For example, one order of selection called first for language, then mathematics, and then science. However, we were unable to follow the protocol in a few schools due to its schedule.

- In every study school, we interviewed the teachers who were observed teaching language, mathematics, and science in grades 7 to 9 in third-cycle schools and grades 10 and 11 in secondary schools.

We interviewed 10 students per school enrolled in grades 9 and 11. We randomly selected these 10 students within their grade from classroom lists obtained before the school data collection visits. When possible, we balanced student selection by gender: 5 male and 5 female students.

Table I.2 shows the sample size of data collected in school in base year. The response rates for the collection of in-school data were more than 95 percent for both the treatment and control groups. For all data collection activities, response rates did not differ meaningfully by study group. Response rates for principals, teachers, students, and classroom observations appear in Appendix B.

[7] Note that, as part of the intervention implemented by FOMILENIO II, a secondary school will be established in each Integrated System in the treatment group. Therefore, at follow-up, we expect a secondary school to be included in all the Integrated Systems of the treatment group.

Table I.2. Sample size of data collected in school in base year

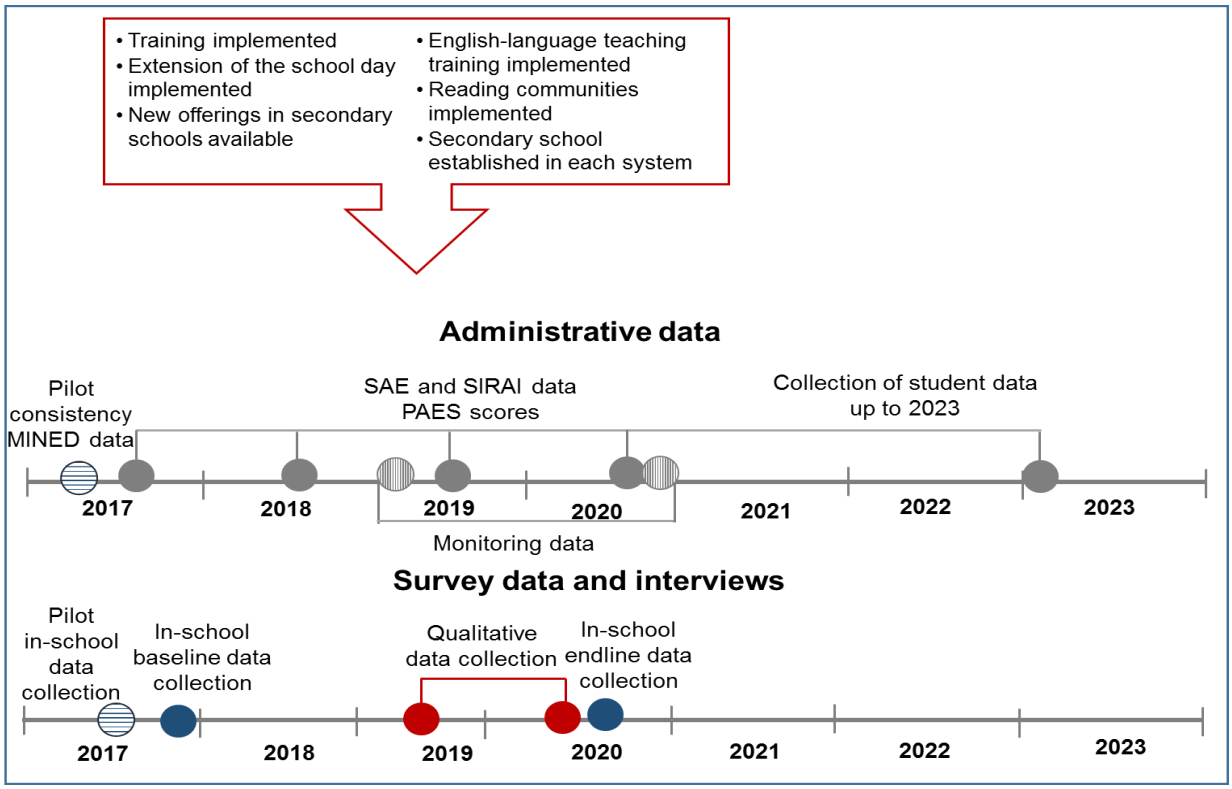
	Third-cycle schools		Secondary schools	
	Treatment group	Control group	Treatment group	Control group
Principals interviewed	54	62	34	42
Classrooms observed	160	185	102	114
Teachers interviewed	146	174	99	115
Students interviewed	539	617	340	400

Source: Principal, teacher, and student surveys and classrooms observed.

Note: The number of teachers surveyed is smaller than the number of classrooms observed because, in schools were the teacher taught more than one subject, we observed the teacher teaching each subject.

Timing of data collection. The timing of the baseline data collection was based on the intervention's implementation time table and other project milestones. According to discussions with FOMILENIO II representatives, key implementation components such as extension of the school day, teacher training, and establishment of secondary schools all started in 2018, and plans call for most schools to be established by the 2019 school year. However, some secondary schools may not undergo improvement until the first trimester of 2020. We planned the data collection schedule around these milestones. We will collect in-school follow-up data at the end of school year 2020 from the data collection sample regardless of schools' implementation progress. In Figure I.3, we present the various milestones, along with the data collection schedule. As shown in the figure, we will collect MINED administrative data up to 2023 enrollment. This follow-up period of educational outcomes will allow us to estimate the impacts of the intervention on dropout until 2022 if the enrollment data for 2023 are available early in the year.

Figure I.3. SI-EITP evaluation data collection schedule, 2017–2023



Note: Extending administrative data collection up to 2024 and the two opportunities for the postsecondary survey (collect data in 2021 (or 2022) or extend it to 2024 (or 2025)) are options that MCC may or may not decide to exercise.

This page has been left blank for double-sided copying.

II. BACKGROUND CHARACTERISTICS OF SCHOOLS, PRINCIPALS, TEACHERS, AND STUDENTS IN THE BASE YEAR

In this chapter, we present the findings from the school, principal, teacher, and student surveys and classroom observations conducted in 2017, the year before the start of implementation of most of the intervention's components.

The results include tests of differences between the treatment and control groups based on regression models that account for the stratification used in the random assignment design (details appear in Appendix B). We found that the groups were balanced on most key characteristics tested, including facilities or services available and teacher and student characteristics. We also found that treatment and control groups were balanced with respect to teacher participation in professional development activities. We did find a few differences between the two groups on some key components of the SI-EITP, such as working together in a pedagogical proposal and extending some hours in third-cycle schools.

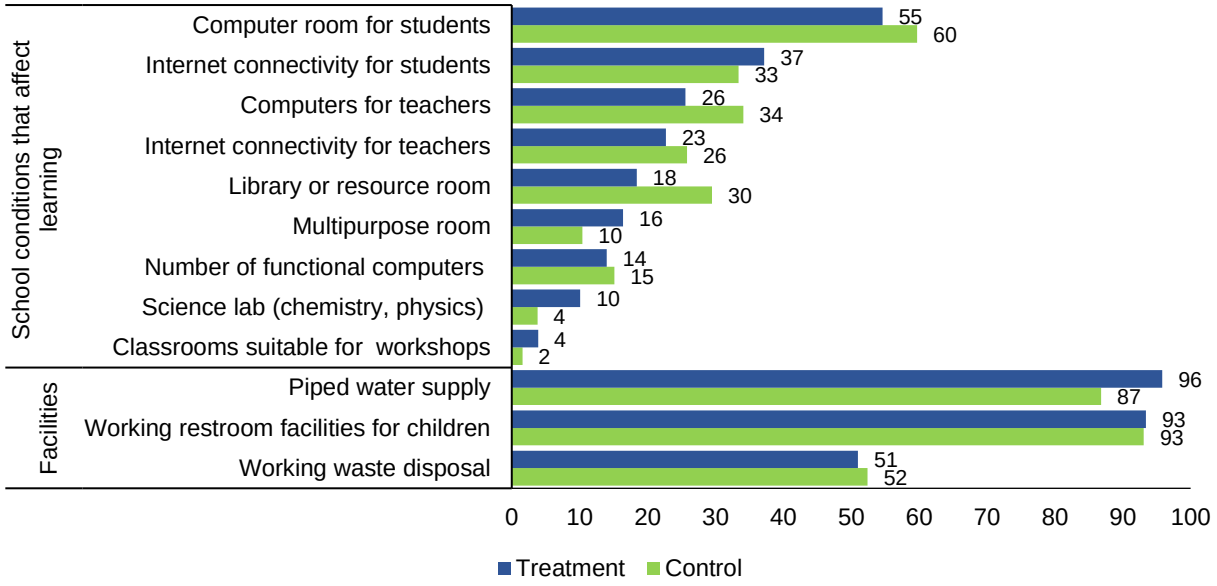
A. Baseline equivalence in school characteristics and infrastructure

1. School infrastructure

Most third-cycle and secondary schools across both the treatment and control groups had access to water and sanitation facilities. Even though the large share of schools in both groups had basic facilities, both groups lacked some basic resources. In both third-cycle and secondary schools, nearly all schools (more than 90 percent in each group) had access to running water and functional toilets. However, only half had a functional waste disposal system. We found a statistically significant difference between treatment and control group schools only with respect to access to running water in secondary schools, with treatment group schools more likely to have piped water compared to control group schools (Figures II.1 and II.2).

Schools conditions that affect learning were relatively similar in both treatment and control group schools. As shown in Figures II.1 and II.2, a little over half of third-cycle schools in both the treatment and control groups had computer rooms for students, yet only a third of the schools had access to the Internet. Information technology (IT) resources for students were even more limited across both third-cycle and secondary schools; on average, schools had 15 functional computers. Moreover, libraries, multipurpose rooms, and science laboratories were not widely available. These findings—which show no differences between treatment and control group schools—are important because IT infrastructure and resources support pedagogy and affect instructional quality. Compared to third-cycle schools, secondary schools' conditions that affect learning were slightly better overall—over 90 percent had computer rooms for students. Teachers in secondary schools also seemed to have more resources; 55 percent had computers and 58 percent had Internet access. We found no differences between treatment and control groups in secondary schools' conditions that affect learning.

Figure II.1. Facilities and utilities available in third-cycle schools in base year (percentages of schools)

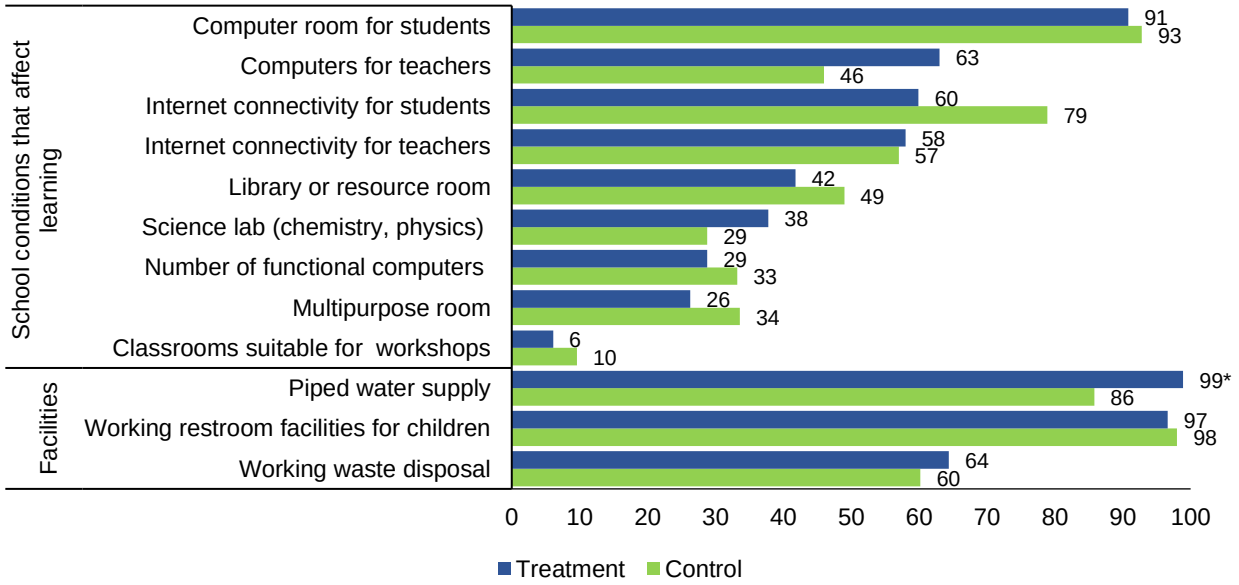


Source: 2017 principal survey.

Note: The bars present group means that are regression-adjusted for the stratification used in the random assignment design.

Number of fuctional computers refers to the average number of computers in a school.

Figure II.2. Facilities and utilities available in secondary schools in base year (percentages unless other unit specified)



Source: 2017 principal survey.

Note: The bars present group means that are regression-adjusted for the stratification used in the random assignment design.

Number of fuctional computers refers to the average number of computers in a school.

* Difference in group means is statistically significant at the .05 level.

2. School shifts and length of school day

There are no significant differences between the treatment and control groups in the number of shifts offered by third-cycle schools, the length of the school day in third-cycle schools, and the class hours per week offered in third-cycle and secondary schools. As shown in Table II.1, schools in the treatment group were more likely than schools in the control group to offer two shifts, but the difference was not statistically significant. In both groups, the duration of the school day in the morning and afternoon shifts in third-cycle schools was approximately 4.5 hours. Furthermore, in third-cycle schools as well as in secondary schools, the average number of total class hours offered per week was similar in both the treatment and control groups (24 hours per week in third-cycle schools and 41 hours per week in secondary schools).

Table II.1. Shifts offered and length of school day in third-cycle schools and class hours per week

	Treatment group (A)	Control group (B)	Difference (A-B)	*p*-value
Third-cycle schools				
School offers third-cycle grades at the following shifts (percentage):				
Morning only	24.7	28.8	-4.1	0.669
Afternoon only	30.5	40.5	-9.9	0.347
Both (morning and afternoon)	44.7	30.7	14.0	0.129
Length of school day (hours per day)				
Morning shift	4.5	4.5	-0.0	0.924
Afternoon shift	4.4	4.4	0.0	0.839
Class hours per week (average)	24.4	24.5	-0.1	0.798
Number of schools	**54**	**62**		
Secondary schools				
Class hours per week (average)	41.2	42.5	-1.3	0.326
Number of schools	**34**	**40**		

Source: 2017 principal survey.

Note: Column A and Column B present group means that are regression-adjusted for the stratification used in the random assignment design. Column A–B presents differences in the regression-adjusted group means between the treatment and control groups. The table includes *p*-values from tests of differences between group means.

3. Participation in the Integrated Systems model, pedagogical proposal, and sharing resources

In both the treatment and control groups, a few principals reported that the school was part of an Integrated System. However, some components of the key activities of the SI-EITP model were more likely to have been implemented in treatment group schools rather than in control group schools. In third-cycle and secondary schools, approximately 12 percent of principals in the treatment group and fewer than 4 percent in the control group reported that the school was part of an Integrated System; the differences were not statistically significant (results not shown). However, when asked about the extension of the school day, which is an

activity promoted by the SI-EITP model, 59 percent of treatment group principals and 39 percent of control group principals reported that the schools offered activities in extended time in the 2017 school year (a statistically significant difference; Table II.2). In addition, in a statistically significant difference, third-cycle schools in the treatment group were more likely than those in the control group to report that they planned to offer activities in extended time in the 2018 school year (28 percentage point difference).

Table II.2. Activities offered in third-cycle schools

	Treatment group (A)	Control group (B)	Difference (A-B)	*p*-value
School offers activities in extended time during 2017 (percent)	58.8	39.4	19.3*	0.049
Number of hours per week in extended time during 2017	2.5	1.6	0.9	0.161
School plans to increase extended time hours in next school year, 2018 (percent)	45.8	18.0	27.9*	0.002
Number of hours per week that are planned for increase in next school year, 2018	3.4	0.9	2.5*	0.006
Number of schools	**34**	**40**		

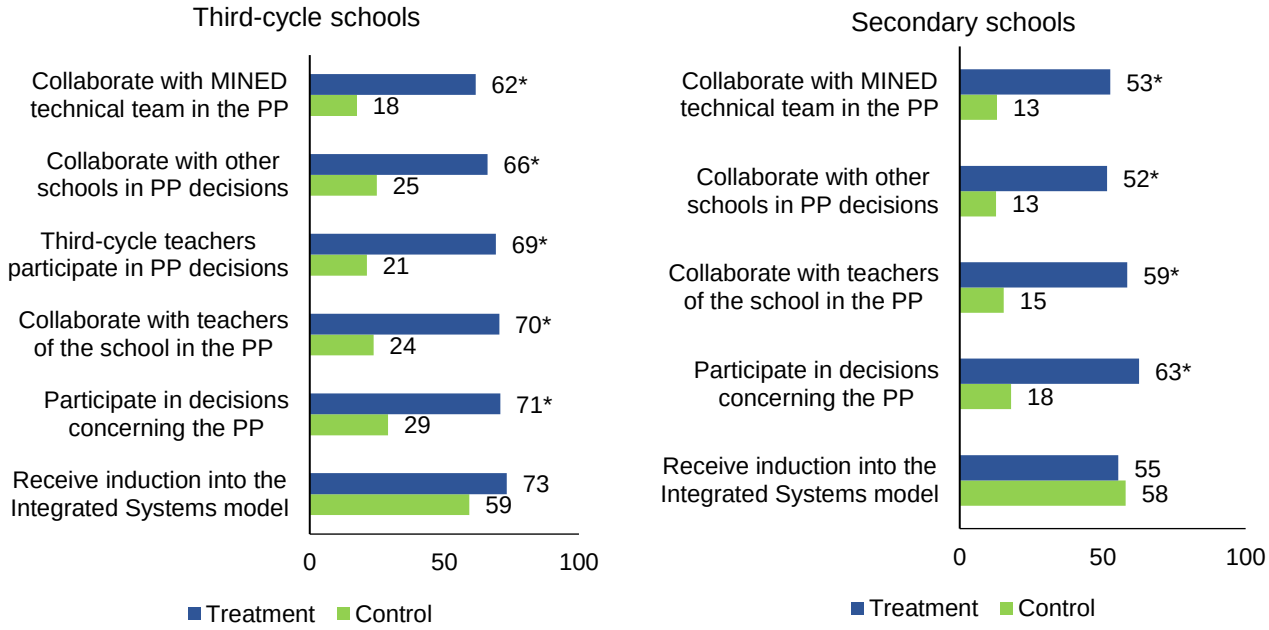
Source: 2017 principal survey.

Note: Column A and Column B present group means that are regression-adjusted for the stratification used in the random assignment design. Column A–B presents differences in the regression-adjusted group means between the treatment and control groups. The table includes *p*-values from tests of differences between group means.

* Difference in group means is statistically significant at the .05 level.

We found no statistically significant differences between the treatment and control groups in the percentage of principals who received induction into the Integrated System model. However, we found that treatment schools were significantly more likely to participate in the development of the pedagogical proposal. In Figure II.3, we show the percentage of teachers who collaborated in development of the pedagogical proposal. Principals of third-cycle and secondary schools in the treatment group were more likely than principals of third-cycle and secondary schools in the control group to participate in decisions concerning the pedagogical proposal and collaborated with MINED's technical team, staff from other schools, and teachers in development of the proposal; the differences were statistically significant.

Figure II.3. Participation in developing the pedagogical proposal (PP)



Source: 2017 principal survey.

Note: The bars present group means that are regression-adjusted for the stratification used in the random assignment design.

* Difference in group means is statistically significant at the .05 level.

We found no statistically significant differences between the treatment and control groups in the percentage of schools that shared resources with nearby schools. Given that an important feature of the SI-EITP model is that schools within an Integrated System will share resources, we asked at baseline about sharing resources. As shown in Table II.3, in third-cycle schools, 63 percent of treatment group schools and 47 percent of control group schools shared resources with nearby schools, but the difference was not statistically significant at the 5 percent level. Treatment group schools were more likely than control group schools to share facilities such as classrooms and libraries, with the difference of 27 percentage points statistically significant. Hence, we concluded that schools in the treatment group started to share resources even without FOMILENIO II's intervention. In secondary schools, 83 percent of treatment group schools and 68 percent of control group schools shared resources with nearby schools, but the difference was not statistically significant.

Table II.3. Percentage of schools that shared resources with nearby schools

	Treatment group (A)	Control group (B)	Difference (A-B)	*p*-value
Third-cycle schools				
Schools shared resources with nearby schools	62.7	47.2	15.5	0.079
Type of resources shared				
Equipment (computers)	46.2	30.2	15.9	0.090
Facilities (classrooms, library, science lab)	45.0	18.4	26.6*	0.001
Pedagogical material (books, pedagogical guides)	13.8	15.8	-2.1	0.779
Other resources	8.6	3.9	4.7	0.308
Number of schools	**54**	**62**		
Secondary schools				
Schools shared resources with nearby schools	83.0	67.9	15.1	0.116
Type of resources shared				
Equipment (computers)	67.5	63.6	3.9	0.727
Facilities (classrooms, library, science lab)	60.7	52.1	8.6	0.481
Pedagogical material (books, pedagogical guides)	11.7	30.6	-18.9	0.060
Other resources	11.5	7.6	3.9	0.608
Number of schools	**34**	**40**		

Source: 2017 principal survey.

Note: Column A and Column B present group means that are regression-adjusted for the stratification used in the random assignment design. Column A–B presents differences in the regression-adjusted group means between the treatment and control groups. The table includes p-values from tests of differences between group means.

* Difference in group means is statistically significant at the .05 level.

B. Baseline equivalence in teachers' and principals' characteristics

1. Teacher characteristics

Teachers had similar demographic characteristics, education, and teaching experience across the treatment and control groups. We did not find statistically significant differences between the treatment and control groups on teacher age or gender. The average teacher was age 43 and had about 17 years of teaching experience, including about 4.5 years in his or her current school (results not shown).

In both the treatment and control groups, approximately half of the teachers taught just one subject, and most teachers reported teaching a subject area related to their training. In Table II.4, we show the percentage of teachers who taught one, two, or three subjects in the study schools. Approximately 50 percent of teachers in third-cycle and secondary schools taught only one subject. Approximately 84 percent of teachers in third-cycle schools in both the treatment and control groups reported teaching a subject related to their training. Similarly, in secondary schools, most teachers reported teaching a subject related to their training. We found no statistically significant differences between the treatment and control groups. The result seemed inconsistent with perceptions of stakeholders who thought that many teachers were teaching a subject in which they were not trained. To understand this inconsistency

more fully, we created another indicator that used two sources of data. First, from the classroom observation, we identified the subject a teacher was teaching during the observation. Then, from the teacher survey, we identified the subjects for which the same teacher reported training. We cross-referenced the two sources of information and determined whether the teacher had received training in the subject taught during the classroom observation (results appear in Table C.1 in Appendix C). The results for teachers in secondary schools were consistent with the teacher reports. However, in third-cycle schools, fewer than 60 percent of teachers reported that they were trained in the subject they taught during the observations. This finding supports the need for the subject-matter teacher training that SI-EITP will offer. The difference between the treatment and control groups was not statistically significant, but it was large; in the treatment group, 60 percent of teachers reported that they trained in the subject taught during the observation, but only 49 in the control group reported the same. When we analyzed these differences by subject area, we found that this difference is driven by teachers observed teaching language. Given that teachers in the treatment group were more likely than teachers in the control group to be trained in language, the difference of 21 percentage points was statistically significant. We did not find statistically significant differences between the treatment and control groups for teachers observed teaching mathematics or science.

Table II.4. Teachers' teaching one or more subjects (percentage)

	Treatment group (A)	Control group (B)	Difference (A-B)	*p*-value
Third-cycle schools				
Teachers teach:				
One subject	49.5	44.1	5.4	0.444
Two subjects	27.9	29.1	-1.3	0.851
Three or more subjects	22.6	26.7	-4.1	0.484
Teachers who report teaching a subject related to their training	85.9	83.8	2.1	0.685
Number of teachers	**146**	**174**		
Secondary schools				
Teachers teach:				
One subject	49.3	52.2	-3.0	0.712
Two subjects	34.4	35.7	-1.4	0.865
Three or more subjects	16.3	12.0	4.3	0.419
Teachers who report teaching a subject related to their training	91.8	87.8	4.0	0.419
Number of teachers	**99**	**115**		

Source: 2017 teacher survey.

Note: Column A and Column B present group means that are regression-adjusted for the stratification used in the random assignment design. Column A–B presents differences in the regression-adjusted group means between the treatment and control groups. The table includes p-values from tests of differences between group means.

2. Teacher-reported participation in professional development in base year

Given that one important component of the SI-EITP model is teacher training, we asked teachers in both the treatment and control groups about their participation in training, who provided the training, and which topics were covered.

Teacher-reported participation in subject-specific training (mathematics, language, biology, chemistry, physics, and social science) was similar in the treatment and control groups in third-cycle and secondary schools. In third-cycle schools, 44 percent of treatment group teachers participated in subject-area training compared to 38 percent of control group teachers. In secondary schools, 51 percent of treatment group teachers participated in subject-area training compared to 59 percent of control group teachers. However, these differences were not statistically significant. In third-cycle and secondary schools, most of the teachers from both the treatment and control groups who participated in training in subject-specific areas indicated that they received the training from MINED. Teachers who participated in technical and technological areas training reported that they received the training from other institutions such as nongovernmental organizations (NGO), telecommunication companies, the Salvadoran Institute of Vocational Training (INSAFORP for its acronym in Spanish), and others. The differences between groups with respect to training providers were not statistically significant (results not shown).

Most teachers had not received any recent training focused on active-learning approaches, gender equality, inclusive classrooms, ICTs, classroom management, soft skills, differentiated instruction, or English-language instruction. In Table II.5, we show the percentage of third-cycle and secondary teachers who received training in topics that are planned to be covered in the SI-EITP training. We found statistically significant differences only in the percentage of third-cycle teachers who reported that they received training in active pedagogies[8] (33 percent of treatment group teachers compared to 20 percent of control group teachers) and the percentage of third-cycle teachers who reported receiving training in inclusive pedagogies (20 percent of treatment group teachers compared to 9 percent of control group teachers). We found no significant differences between groups in secondary schools in the types of training received. In third-cycle and secondary schools, the majority of the teachers from both the treatment and groups reported that they received training in gender equity, classroom management, and soft skills from institutions—from, for example, NGOs, international organizations, and INSAFORP. Teachers who participated in training in active-learning pedagogies indicated that they received the training from MINED or other institutions (mainly NGOs). We found no significant differences between groups with respect to the providers that delivered the training.

[8] When teachers were asked about their participation in active pedagogies training we did not specify a definition of active pedagogies. Since there is no standard definition of active pedagogies, teachers might refer to different teaching methods. The active instructional practices coded with the Stallings method is one of many ways to define what active pedagogies are. Therefore, the information collected from these two sources does not represent the same information.

Table II.5. Teachers' participation in reading instruction professional development in base year

	Treatment group (A)	Control group (B)	Difference (A-B)	*p*-value
Third-cycle schools				
Percentage of teachers who reported participation in professional development focused on:				
Subject-specific areas	44.2	38.0	6.2	0.303
Active-learning pedagogies	33.4	19.9	13.6*	0.017
Gender equity	26.3	18.4	8.0	0.102
Technical and technological areas	21.2	23.7	-2.5	0.714
Inclusive classrooms or educational inclusion	19.8	8.6	11.1*	0.002
Class management, discipline, or positive classroom climate	13.7	10.8	2.9	0.452
Life and work skills (soft skills)	9.6	4.6	5.0	0.071
Differentiated instruction	7.8	4.0	3.8	0.126
English language	2.8	1.9	1.0	0.472
Other	4.8	3.2	1.5	0.531
Number of teachers	**146**	**174**		
Secondary schools				
Percentage of teachers who reported participation in professional development focused on:				
Subject-specific areas	51.3	59.1	-7.7	0.347
Active-learning pedagogies	29.5	33.4	-3.9	0.641
Technical and technological areas	24.4	31.8	-7.5	0.347
Inclusive classrooms or educational inclusion	17.3	16.8	0.4	0.934
Gender equity	17.0	20.5	-3.5	0.495
Class management, discipline, or positive classroom climate	16.7	12.3	4.4	0.438
Differentiated instruction	13.2	8.0	5.2	0.292
Life and work skills (soft skills)	10.6	5.2	5.4	0.191
English language	2.3	0.1	2.2	0.152
Other	4.8	1.6	3.2	0.182
Number of teachers	**99**	**115**		

Source: 2017 teacher survey.

Note: Column A and Column B present group means that are regression-adjusted for the stratification used in the random assignment design. Column A–B presents differences in the regression-adjusted group means between the treatment and control groups. The table includes p-values from tests of differences between group means.

* Difference in group means is statistically significant at the .05 level.

3. Teachers' and principals' perceptions about parent and community school involvement

We found no significant differences in the percentage of teachers who reported that at least 50 percent of the parents of students in their classrooms were involved in school activities. Teachers and principals reported the percentage of parents who, according to their estimates, were involved in school activities. In third-cycle and secondary schools, most teachers and principals reported that at least 50 percent of parents attended meetings at the request of

teachers, and more than 80 percent reported that at least half of parents were aware of teachers' or principals' decisions, students' academic performance, and students' discipline. Fewer than 60 percent of teachers and principals reported that, in both the treatment and control groups, at least half of parents help students with their homework. We found no statistically significant differences between the treatment and control groups in parent school involvement as reported by teachers and principals. See Table II.6.

Table II.6. Teachers' and principals' perception of parent and community involvement in base year

	Treatment group (A)	Control group (B)	Difference (A-B)	*p*-value
Third-cycle schools				
Attend meetings at teacher request	92.8	94.9	-2.1	0.416
Aware of teachers' or principals' decisions	88.8	93.1	-4.3	0.177
Aware of students' academic performance	88.5	82.2	6.3	0.096
Aware of students' behavior at school	88.1	85.5	2.6	0.473
Help with food preparation for students	84.7	88.4	-3.8	0.284
Attend cultural or athletic activities at school	82.3	78.2	4.1	0.401
Communicate with teachers regularly	70.6	65.7	4.9	0.314
Volunteer in school activities	65.2	64.0	1.2	0.829
Help the student with homework	59.4	53.2	6.3	0.193
Number of teachers and principals	**200**	**236**		
Secondary schools				
Attend meetings at teacher request	95.5	95.1	0.4	0.884
Aware of teachers' or principals' decisions	92.8	90.0	2.8	0.335
Aware of students' academic performance	87.9	86.1	1.8	0.665
Aware of students' behavior at school	87.4	85.5	1.9	0.632
Help with food preparation for students	64.5	71.0	-6.5	0.43
Attend cultural or athletic activities at school	60.9	72.3	-11.5	0.059
Volunteer in school activities	59.6	61.5	-1.9	0.75
Communicate with the teachers regularly	53.6	65.8	-12.2	0.052
Help the student with homework	51.9	57.0	-5.1	0.379
Number of teachers and principals	**133**	**155**		

Source: 2017 teacher and principal surveys.

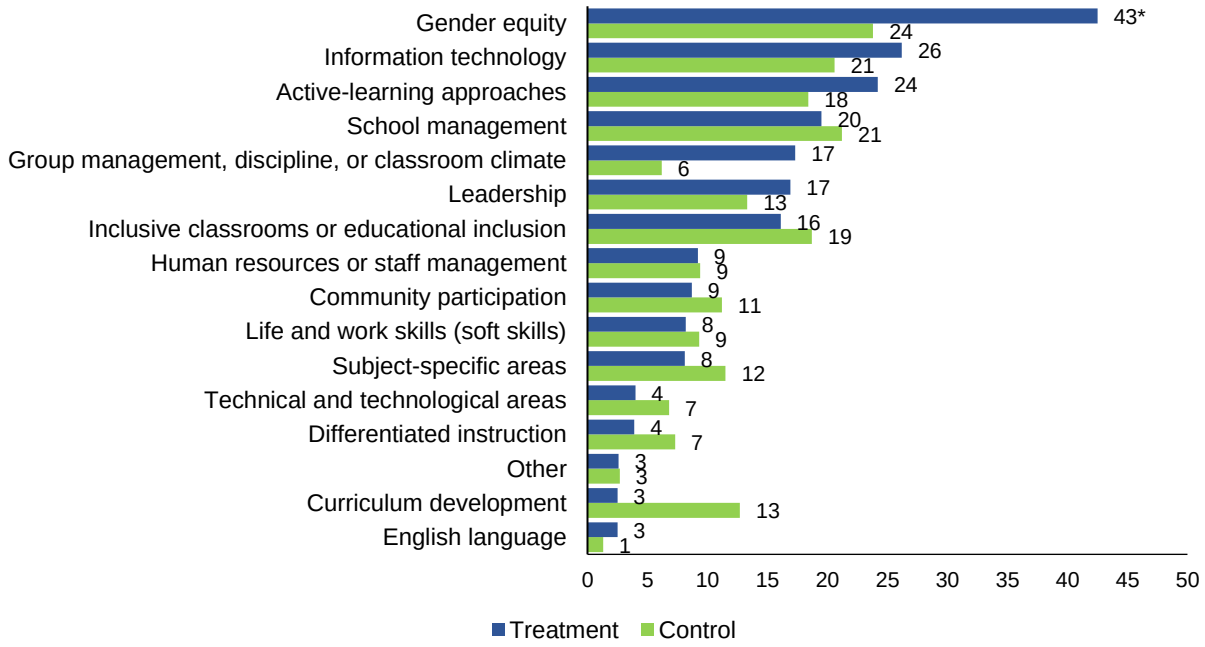
Note: Column A and Column B present group means that are regression-adjusted for the stratification used in the random assignment design. Column A–B presents differences in the regression-adjusted group means between the treatment and control groups. The table includes p-values from tests of differences between group means.

4. Principal-reported participation in professional development in base year

In third-cycle schools, we found—with one exception—no statistically significant differences between the treatment and control groups with respect to the professional development received by principals. In both the treatment and control groups, roughly one in

four principals had received training in information technology, active pedagogies, and school management (Figure II.6). We found only one statistically significant difference in the types of training received. For treatment group schools, 43 percent of principals had received training in gender equity compared to 24 percent of principals in control group schools.

Figure II.6. Principal-reported participation in professional development in base year in third-cycle schools



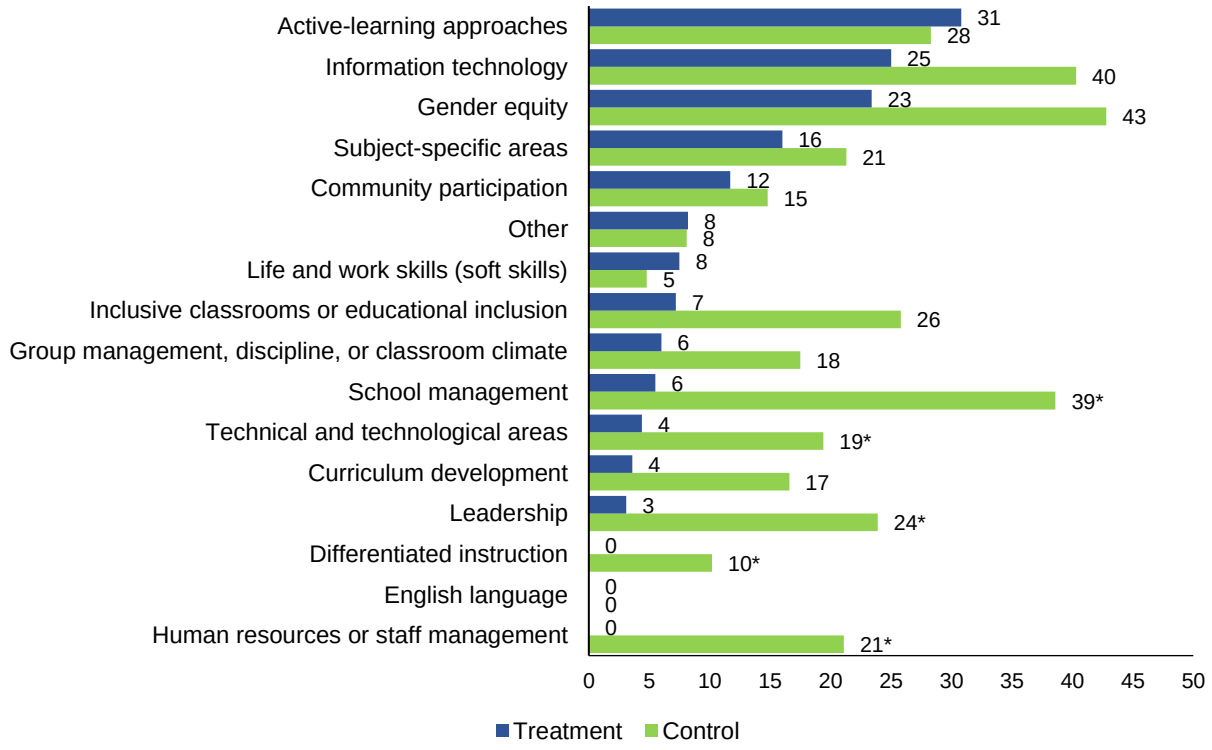
Source: 2017 principal survey.

Note: The bars present group means that are regression-adjusted for the stratification used in the random assignment design.

* Difference in group means is statistically significant at the .05 level.

In secondary schools, more principals in the control group than in the treatment group received training in several areas. In Figure II.7, we show that more principals in the control group versus the treatment group reported that they received training in school management, technological areas, leadership, differentiated instruction, and staff management as compared to principals in the treatment group; these differences between study groups were statistically significant. This initial difference, may have provided an advantage to the control group during the baseline year. However, given that FOMILENIO II's implementation of the SI-EITP model includes intensive training for principals and school committees, we would expect the treatment group to receive more training than the control group in the coming school years.

Figure II.7. Principal-reported participation in professional development in base year in secondary schools



Source: 2017 principal survey.

Note: The bars present group means that are regression-adjusted for the stratification used in the random assignment design.

* Difference in group means is statistically significant at the .05 level.

C. Baseline equivalence in student characteristics

We found statistically significant differences between the treatment and control groups in students' personal characteristics (gender and age), but not in student-reported access to school (means of transportation) or commute time. In third-cycle schools, less male students in the treatment group (49.7) were interviewed than in the control group (50.1). In secondary schools, more male students in the treatment group were interviewed (50.7) than in the control group (47.9). These differences were statistically significant but small. It is important to note that the student selection protocol called for 5 male and 5 female students to be selected when possible. Therefore, the differences were attributable to the inability to find enough male or female students in the classroom to be interviewed. In third-cycle schools, students in the control group were slightly younger than those in the treatment group. The mean age for treatment group students was 16.1 years compared with 15.8 years for control group students—a three-month difference. This difference was statistically significant. In both the treatment and control groups, approximately 70 percent of third-cycle students walked to school, and 9 percent used public transportation, with an average commute time to school of 17 minutes. We found no statistically significant differences between the treatment and control groups in terms of means of transport and commute time. In secondary schools, almost half of students walked to school, and 30 percent used public transportation. We found no statistically significant differences in

commuting time between the treatment and control groups; on average, the commuting time to secondary schools was 20 minutes for both groups. Table II.7.

Table II.7. Student background characteristics in base year

	Treatment group (A)	Control group (B)	Difference (A-B)	*p*-value
Third-cycle schools				
Male students interviewed (percentage)	49.7	50.1	-0.5*	0.045
Age (average)	16.1	15.8	0.3*	0.001
Transportation to go to school				
Walk	74.8	73.6	1.1	0.830
Public transportation	9.9	10.9	-1.0	0.741
Motorcycle	2.9	0.8	2.1	0.106
Bicycle	5.4	6.4	-1.0	0.708
Car or school bus	2.6	5.7	-3.1	0.103
Other	-0.0	0.6	-0.6	0.056
Two means of transportation	3.9	1.7	2.1	0.134
Commuting time to school (average)	16.9	16.8	0.1	0.931
Number of students	**539**	**617**		
Secondary schools				
Male students interviewed (percentage)	50.7	47.9	2.8*	0.029
Age (average)	18.0	18.1	-0.1	0.256
Transportation to go to school				
Walk	51.2	54.1	-2.9	0.648
Public transportation	29.1	29.4	-0.3	0.960
Motorcycle	5.5	1.4	4.0	0.139
Bicycle	3.7	3.2	0.5	0.850
Car or school bus	4.2	6.6	-2.5	0.195
Other	0.9	0.7	0.2	0.850
Two means of transportation	5.4	4.4	1.0	0.561
Commuting time to school (average)	20.1	19.8	0.3	0.854
Number of students	**340**	**400**		

Source: 2017 student survey.

Note: Column A and Column B present group means that are regression-adjusted for the stratification used in the random assignment design. Column A–B presents differences in the regression-adjusted group means between the treatment and control groups. The table includes p-values from tests of differences between group means.

* Difference in group means is statistically significant at the .05 level.

This page has been left blank for double-sided copying.

III. TEACHING PRACTICES AND USE OF CLASS TIME IN THE BASE YEAR

We obtained data on teacher activities and practices from two sources, the teacher survey and the Stallings classroom observation. These two sources offer complementary information on what happens in the classrooms.

- The teacher survey allows to obtain information on the frequency *during a month* in which teachers report conducting common teacher practices, for example debate, experimentation, paired or individual problem solving (see Figure III.1).
- The teacher observation conducted with the Stallings protocol allows to obtain information on how much *class time* the teacher spent conducting core pedagogical practices, managing the class, or off-task. As discussed in Chapter I, the Stallings protocol specifies six core pedagogical practices for example reading aloud, demonstration, or seatwork (see Table III.1 for the full list).

This section presents teacher practices results from both data sources to provide a more complete description of teachers' activities. However, we discuss the results separately because these two sources provide information on different aspects of teaching practices and activities. The survey provides teacher perceptions on frequency during a month and the Stallings provides observed frequency of practices during a class. More importantly, the list of teachers' practices used in the teacher survey does not directly overlap with the list of teacher practices in Stallings observation.

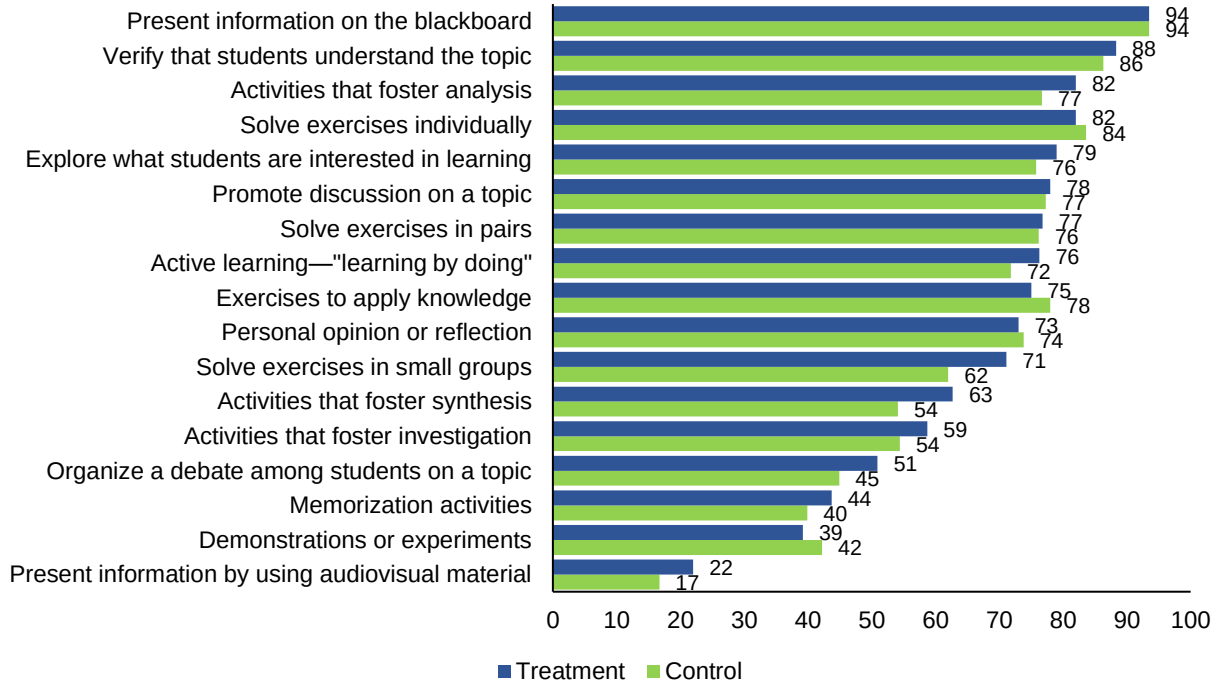
We first discuss teacher practices findings from the teacher survey. The teacher baseline survey used includes a list of general practices commonly conducted by teachers because, at the time of baseline data collection, the implementation was still developing and specific target teacher practices had not been identified. However, the teacher survey will be modified at follow-up to incorporate targeted teacher practices aligned with the current implementation of the SI-EITP model.

Then, we discuss extensively the results on teachers' use of class time obtained from the Stallings observation. Time on task is a key outcome for the evaluation, given that FOMILENIO's implementation of SI-EITP aims to increase time in academics activities. Other indicators obtained from the Stallings and discussed here are materials used in class and student engagement.

A. Baseline equivalence of self-reported teaching practices

In third-cycle and secondary, teachers in the treatment and control groups report performing common instructional practices with similar frequency. In the teacher survey, we asked teachers to report the frequency with which they performed some general teaching practices such as presenting information on the blackboard, solving exercises in pairs or small groups, and promoting discussion. For both third-cycle and secondary schools, we did not find statistically significant differences in teachers' reported likelihood of performing these teaching practices more than once per week (Figures III.1 and III.2).

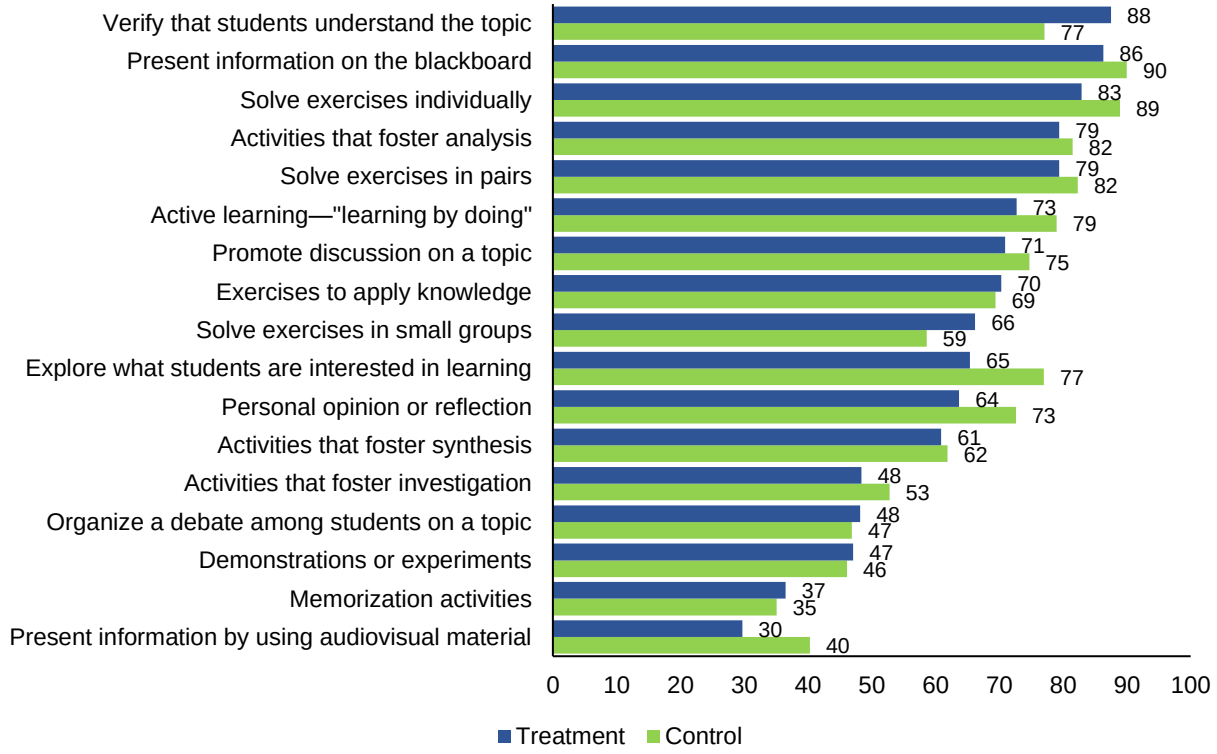
Figure III.1. Teaching practices performed more than once per week (third-cycle teachers' report)



Source: 2017 Teacher survey.

Note: The bars present group means that are regression-adjusted for the stratification used in the random assignment design.

Figure III.2. Teaching practices used more than once per week (secondary teachers' report)



Source: 2017 Teacher survey.

Note: The bars present group means that are regression-adjusted for the stratification used in the random assignment design.

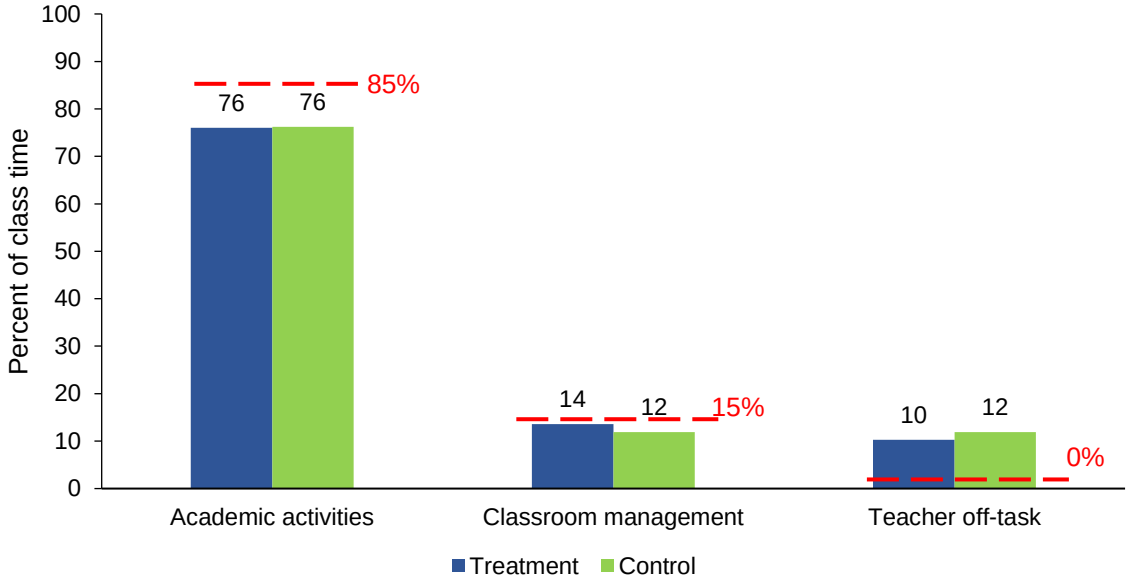
B. Baseline equivalence of observed third-cycle teachers' use of class time

Teachers in the treatment and control groups were observed using class time in similar ways, spending the large share of their time on academic activities and less time on class management. However, both groups also spent some time off-task. In the class period we observed, teachers in both the treatment and control groups spent approximately 76 percent of their time on academic activities, 13 percent on classroom management, and 11 percent off-task. Compared to the international benchmarks[9] discussed in Chapter I, teachers in both the treatment and control groups spent almost 10 percentage points less time on academic activities than the recommended best practice guidelines (76 percent versus the recommended 85 percent). Furthermore, teachers in both groups spent an average of about 11 percent of class time off-task

[9] In the benchmark defined by Stallings and Knight (2003), teachers spend an average of 85 percent of class time on instruction, 50 percent on "active" instructional activities, and not more than 35 percent of total class time on "passive" instruction. Teachers also spend 15 percent or less of class time on organizing and managing. These teachers were never observed off-task.

when the best practice guidelines recommend no off-task class time at all (Figure III.3). The differences between groups in class time use were not statistically significant.

Figure III.3. Teachers' use of class time in third-cycle schools



Source: 2017 classroom observation.

Note: The bars present group means that are regression-adjusted for the stratification used in the random assignment design.

Teachers in the treatment and control groups were observed distributing class time in academic instruction in similar ways across the core pedagogical practices. In the class observed, teachers in both the treatment and control groups spent approximately 50 percent of their time on active instruction activities and about 27 percent on passive instruction (Table III.1). The total time spent on active instruction practices is consistent with the recommended amount—50 percent of class time. The teachers in both groups spent similar amounts of time on the four active instruction practices (read aloud, exposition or demonstration, questions and answers or discussion, and practice and drill), but we found a statistically significant difference in one passive instruction practice. Treatment group teachers on average devoted a greater proportion of class time to monitoring seatwork (treatment group 10 percent and control group 7 percent).

Table III.1. Time spent on core pedagogical practices

	Treatment group (A)	Control group (B)	Difference (A-B)	*p*-value
Active instruction	50.3	48.6	1.7	0.550
Reading aloud	1.6	2.0	-0.4	0.551
Exposition or demonstration	30.3	29.2	1.0	0.677
Question and answer/discussion	18.2	17.5	0.8	0.593
Practice and drill	0.3	0.0	0.2	0.072
Passive instruction	25.7	27.6	-1.8	0.462

	Treatment group (A)	Control group (B)	Difference (A-B)	*p*-value
Seatwork	10.0	7.0	2.8*	0.031
Copying	16.0	20.0	-4.7	0.056
Number of classrooms	**160**	**185**		

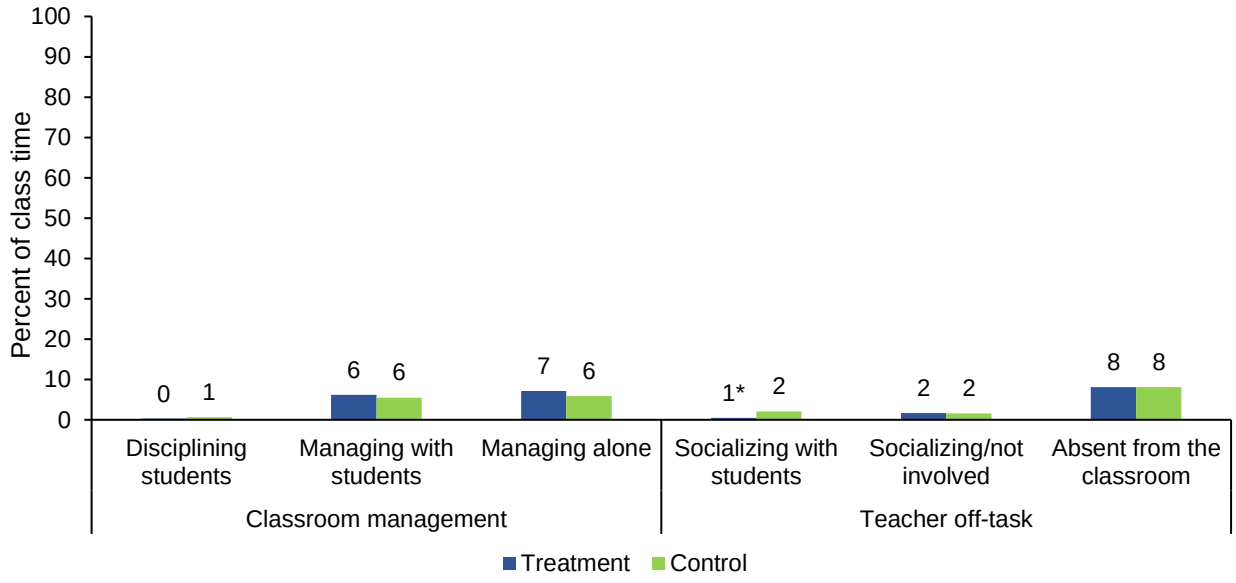
Source: 2017 classroom observation.

Note: Column A and Column B present group means that are regression-adjusted for the stratification used in the random assignment design. Column A–B presents differences in the regression-adjusted group means between the treatment and control groups. The table includes p-values from tests of differences between group means.

* Difference in group means is statistically significant at the .05 level.

Teachers in both treatment and control groups were observed spending a similar proportion of class time in classroom management activities. In both groups, the large proportion of teachers' off-task time reflected time that teachers spent outside the classroom, particularly at the beginning of class. We found a statistically significant difference in the time teachers spent socializing with students. In Figure III.4, we show the breakdown of class time spent on class management and off-task activities among teachers in both the treatment and control groups. We found no statistically significant differences in the time spent on each of the classroom management activities. Teachers in the two groups spent most of the time managing the class alone or with the help of students. The breakdown of off-task time shows that teachers in both groups were absent from the classroom 8 percent of total class time. Further, teachers' off-task time usually occurred at the beginning of class. We estimated the percentage of time that teachers were absent from the room in each instance observed and found that, in both groups, half of teachers' time out of the classroom was recorded during the first two snapshots (4 percent of the time). We present detailed information on use of class time in Appendix C. We also found that teachers in the control group versus the treatment group spent, on average, significantly more time socializing with students (1 percentage point difference).

Figure III.4. Breakdown of teacher time managing the classroom and off-task



Source: 2017 classroom observation.

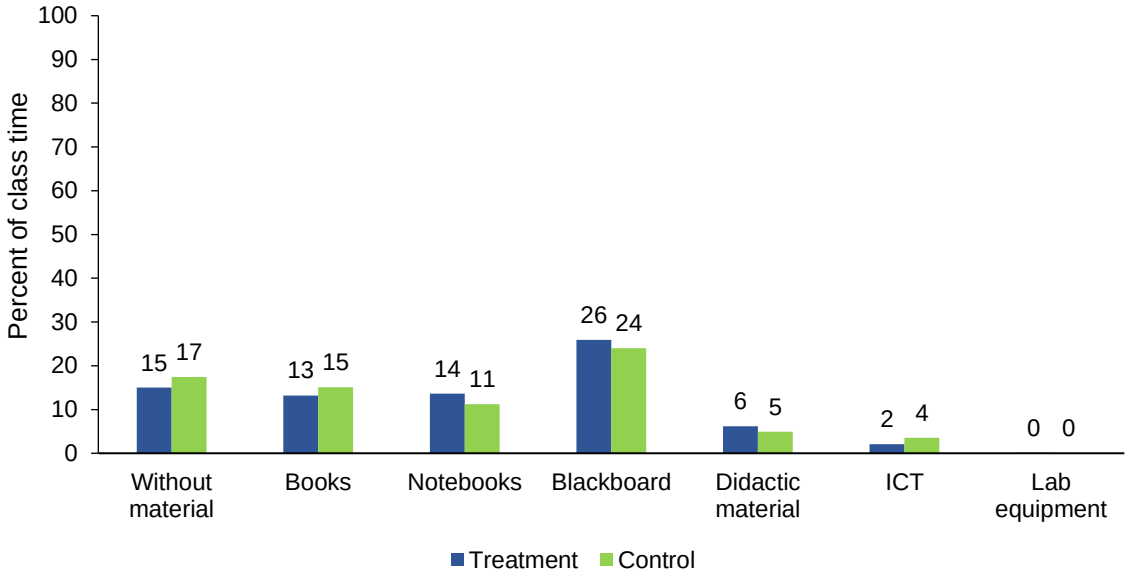
Note: The bars present group means that are regression-adjusted for the stratification used in the random assignment design.

* Difference in group means is statistically significant at the .05 level.

The Stallings observation instrument also allows us to look at teachers' use of materials during class time.

In the class observed, teachers in the treatment and control groups used learning materials in similar ways, generally relying on the blackboard or lecturing without any material most of the class time. Teachers in both the treatment and control groups spent roughly 25 percent of their instruction time using the blackboard. Treatment and control group teachers spent 16 percent of class time teaching without any material, 14 percent of class time using books, and 13 percent using notebooks. In both groups, teachers made little use of ICT or laboratory material (Figure III.5).

Figure III.5 Teachers' use of learning material

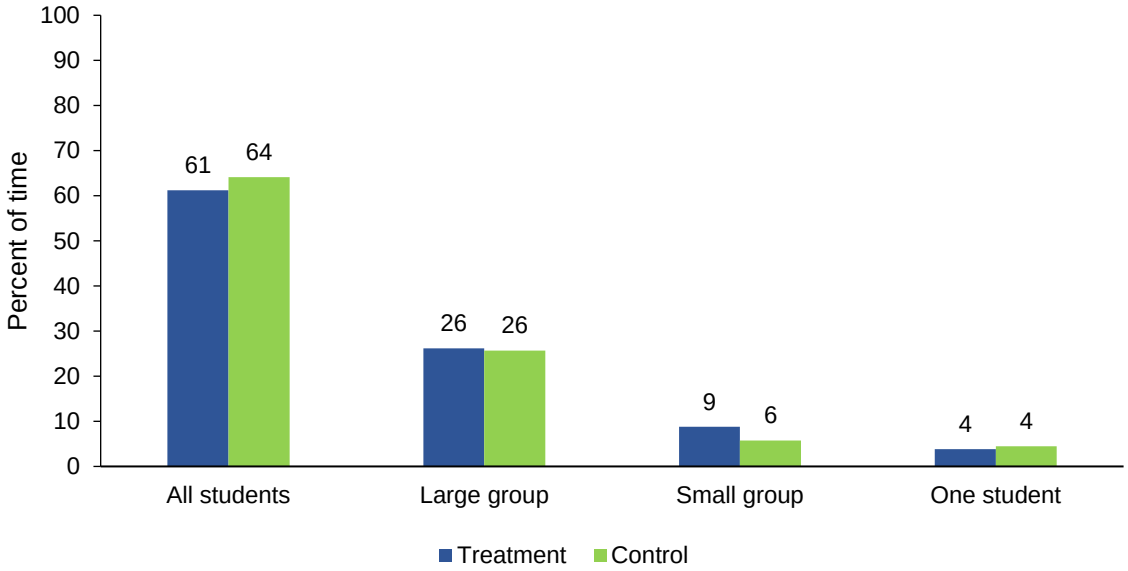


Source: 2017 classroom observation.

Note: The bars present group means that are regression-adjusted for the stratification used in the random assignment design.

In both the treatment and control groups, when the teacher was performing an academic activity, a large group of students or all students were engaged with the teacher most of that time. We estimated the percentage of time that one student, a large group or small group of students, or all students were engaged with the teacher when she or he was performing an academic activity (Figure III.6). We found no significant differences between the treatment and control groups in the percentage of time that groups of students were engaged with the teacher when she or he was performing an academic activity. Approximately 62 percent of the time when a teacher was performing an academic activity, all students were engaged, and 26 percent of the time a large group was engaged.

Figure III.6. Percentage of time students were engaged with the teacher when the teacher was performing an academic activity



Source: 2017 classroom observation.
Note: The bars present group means that are regression-adjusted for the stratification used in the random assignment design.
This outcome is expressed as a percentage of the total time the teacher was engaged in instructional activities.

We found no statistically significant differences between the treatment and control groups in the percentage of time that students were off-task when their teacher was performing an academic activity. In Table III.2, we present the percentage of time that at least one student was off-task when the teacher was performing an academic activity and the percentage of time that a large group of students (six or more) was off-task when the teacher was performing an academic activity. In both the treatment and control groups, at least one student was off-task for approximately 30 percent of the time when a teacher performed an instructional activity. However, large groups of students were off-task for only 6 percent of the time when the teacher was leading an academic activity.

Table III.2. Percentage of time students were off-task when their teacher was performing an academic activity

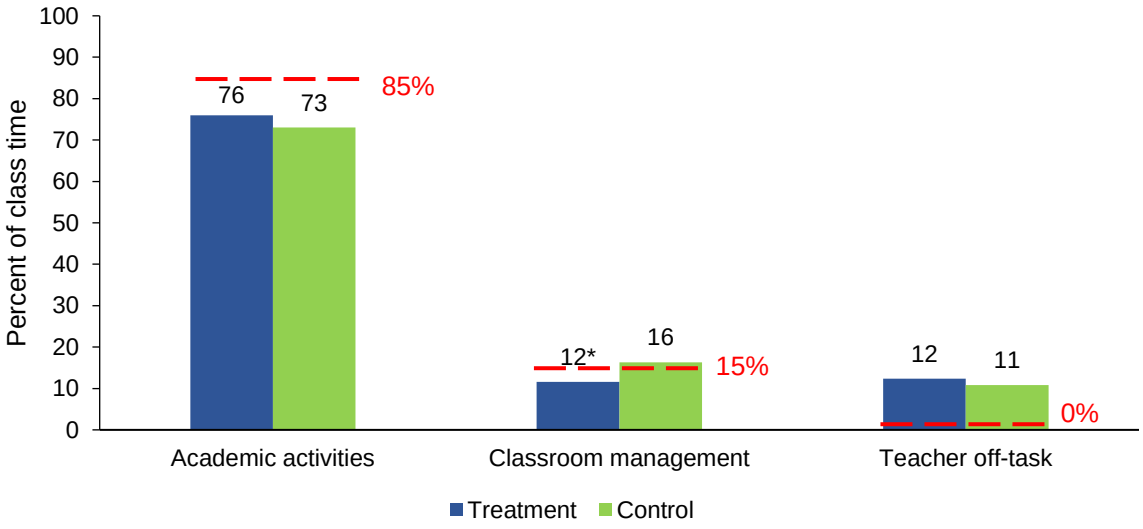
	Treatment group (A)	Control group (B)	Difference (A-B)	*p*-value
At least one student was off-task when the teacher was performing an academic activity	30.0	30.0	0.0	0.999
Large group (at least six students) was off-task when the teacher was performing an academic activity	6.5	5.2	1.3	0.485
Number of classrooms	**160**	**185**		

Source: 2017 classroom observation.
Note: Column A and Column B present group means that are regression-adjusted for the stratification used in the random assignment design. Column A–B presents differences in the regression-adjusted group means between the treatment and control groups. The table includes p-values from tests of differences between group means.
*Difference in group means is statistically significant at the .05 level.

C. Baseline equivalence of teachers' use of class time in secondary schools

Secondary school teachers in the treatment and control groups spent a similar percentage of class time on academic activities and off-task. But we found statistically significant differences in the percentage of class time they spent managing the class. Teachers in the treatment and control groups spent on average of 73 to 76 percent of class time, respectively, on academic activities and approximately 11 percent of time off-task. We also found that teachers in the control group spent a significantly greater proportion of class time on classroom management activities than did teachers in the treatment group (12 percent in the treatment group and 16 percent in the control group). Compared to the best-practice benchmark, teachers in secondary schools spent about 10 percentage points less than the recommended 85 percent of time on academic activities as well as more time off-task than the recommended no time off-task for effective teaching (Figure III.7).

Figure III.7. Teachers' use of class time in secondary schools



Source: 2017 classroom observation.

Note: The bars present group means that are regression-adjusted for the stratification used in the random assignment design.

* Difference in group means is statistically significant at the .05 level.

Teachers in the treatment and control groups divided class time in academic instruction in similar ways across the core pedagogical practices. The share of class time that teachers spent on active and passive instruction was consistent with the recommended amount. Teachers in both groups spent 52 percent of class time on active instruction activities and 23 percent on passive instruction. The groups spent similar amounts on the four active instruction practices (read aloud, exposition or demonstration, questions and answers or discussion, and practice and drill). They also spent a similar share of class time on the two passive instruction practices (monitoring seatwork and copying) (Table III.3).

Table III.3. Proportion of class time spent on core pedagogical practices

	Treatment group (A)	Control group (B)	Difference (A-B)	*p*-value
Active instruction	52.4	51.8	0.7	0.853
Reading aloud	2.7	4.0	-1.3	0.419
Exposition or demonstration	31.6	28.7	2.9	0.406
Question and answer/discussion	17.9	18.9	-1.0	0.66
Practice and drill	0.3	0.1	0.2	0.585
Passive instruction	23.6	21.2	2.4	0.388
Seatwork	12.1	10.0	2.1	0.357
Copying	11.5	11.2	0.3	0.875
Number of classrooms	**102**	**114**		

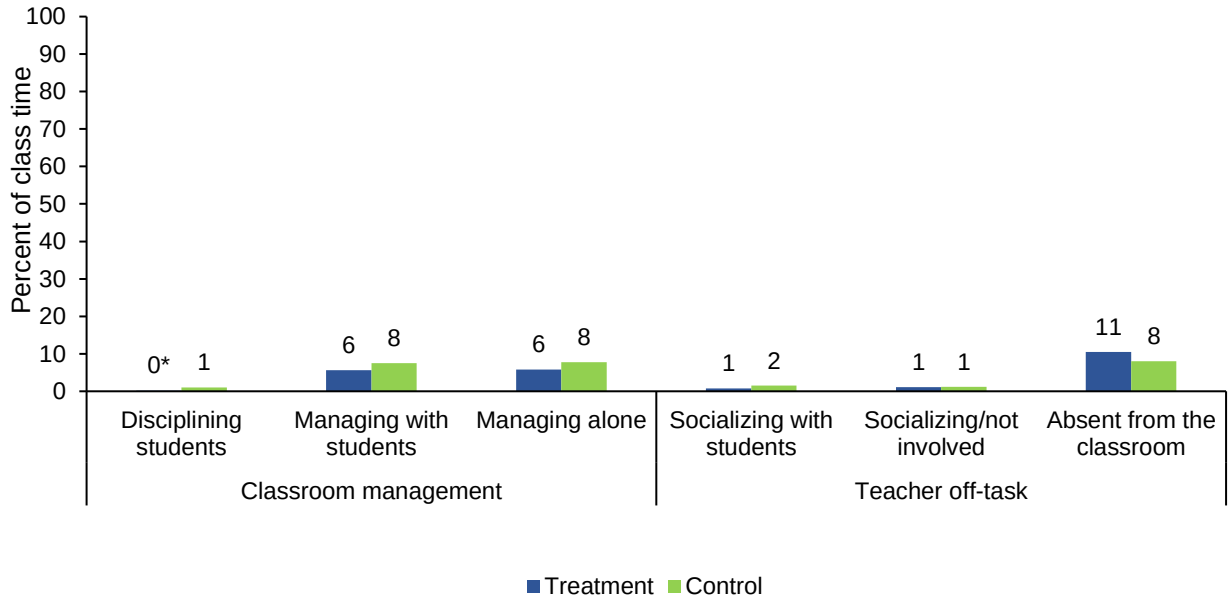
Source: 2017 classroom observation.

Note: Column A and Column B present group means that are regression-adjusted for the stratification used in the random assignment design. Column A–B presents differences in the regression-adjusted group means between the treatment and control groups. The table includes p-values from tests of differences between group means.

* Difference in group means is statistically significant at the .05 level.

We found a significant difference in the proportion of class time teachers spent disciplining students. In both the treatment and control groups, the large share of teachers' off-task time reflected teachers' time outside the classroom, particularly at the beginning of class. In Figure III.6, we show the breakdown of class time spent on class management. Teachers in both groups spent most of the time managing the class alone or with the help of the students, with no significant differences between the treatment and control groups. However, we found a statistically significant difference in the share of class time teachers devoted to disciplining students; teachers in the control group spend on average 1 percent of class time disciplining students and the treatment group spends no time on this activity**.** In Figure III.8, we also show how teachers spent their off-task time: teachers in the treatment group were absent from the classroom 11 percent of the time compared with 8 percent of teachers in the control group. This difference was not statistically significant. Most of teacher absences from the classroom usually occurred at the beginning of the class. In the treatment group, teachers spent 6 percent (out of the total 11 percent of time absent) absent from the classroom during the first two instances observed. Teachers in the control group spent 3 percent (out of the 8 percent of time absent) absent from the classroom during the first two instances observed. Detailed results appear in Appendix C.

Figure III.8. Breakdown of teacher time managing the classroom and off-task



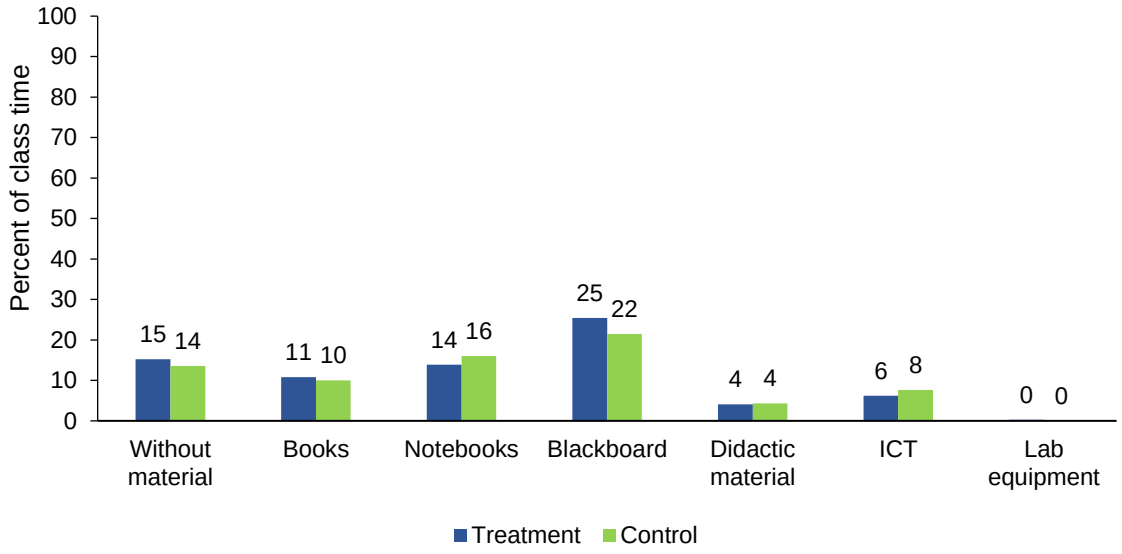
Source: 2017 classroom observation.

Note: The bars present group means that are regression-adjusted for the stratification used in the random assignment design.

* Difference in group means is statistically significant at the .05 level.

Teachers in the treatment and control groups divided the in-class time similarly with respect to the use of learning material. Teachers in the treatment and control groups spent roughly 24 percent of their instruction time using the blackboard and 13 percent of their instruction time teaching without any material. Teachers also spent 10 percent of their instruction time using books and 15 percent using notebooks. In both groups, teachers made little use of ICT or laboratory material. We found no differences between groups in the share of time in which they used learning material (Figure III.9).

Figure III.9. Teachers' use of learning material in secondary schools

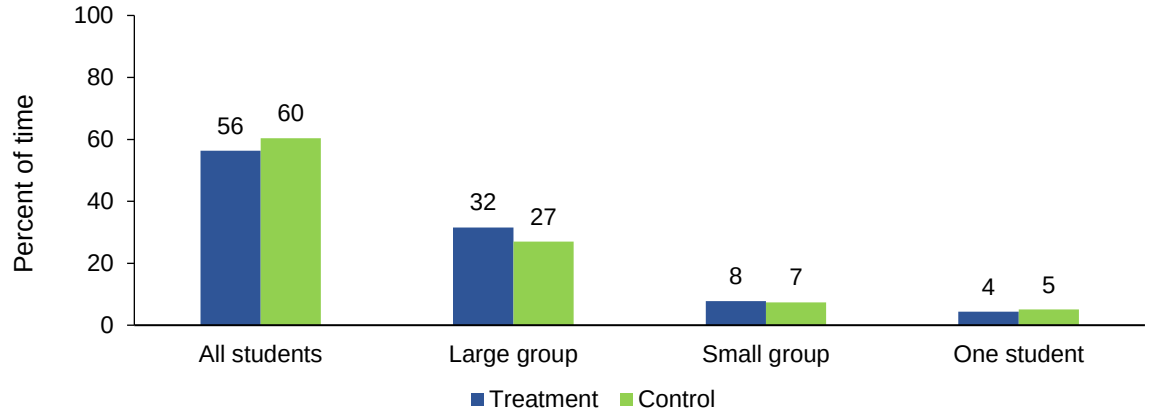


Source: 2017 classroom observation.

Note: The bars present group means that are regression-adjusted for the stratification used in the random assignment design.

In the treatment and control groups, when the teacher was performing an academic activity, large groups of students or most students were engaged in the academic activity for most of the time. We found no significant differences in the share of time students were engaged when the teacher was performing an academic activity. All students were engaged for 58 percent of the time that the teacher was teaching, and large groups were engaged in the teacher's academic activity for 29 percent of the time (Figure III.10).

Figure III.10. Percentage of time students were engaged when the teacher was performing an academic activity



Source: 2017 classroom observation.

Note: The bars present group means that are regression-adjusted for the stratification used in the random assignment design.

This outcome is expressed as a percentage of the time the teacher was engaged in instructional activities.

We found no significant differences between the treatment and control groups in the share of time that students were off-task when the teacher was performing an academic activity. We looked at the number of students off-task when the teacher was performing an academic activity. In both the treatment and control groups, at least one student was off-task 33 percent of the time when the teacher was performing an academic activity. However, a large group of students was off-task only 6 percent of the time when the teacher was performing an academic activity (Table III.4).

Table III.4. Percentage of time students were off-task when their teacher was performing an academic activity

	Treatment group (A)	Control group (B)	Difference (A-B)	*p*-value
At least one student was off-task when the teacher was performing an academic activity	34.9	31.4	3.5	0.420
Large group (at least six students) was off-task when the teacher was performing an academic activity	5.8	6.4	-0.6	0.793
Number of classrooms	**102**	**114**		

Source: 2017 classroom observation.

Note: Column A and Column B present group means that are regression-adjusted for the stratification used in the random assignment design. Column A–B presents differences in the regression-adjusted group means between the treatment and control groups. The table includes p-values from tests of differences between group means.

This outcome is expressed as a percentage of the total time the teacher was engaged in instructional activities.

* Difference in group means is statistically significant at the .05 level.

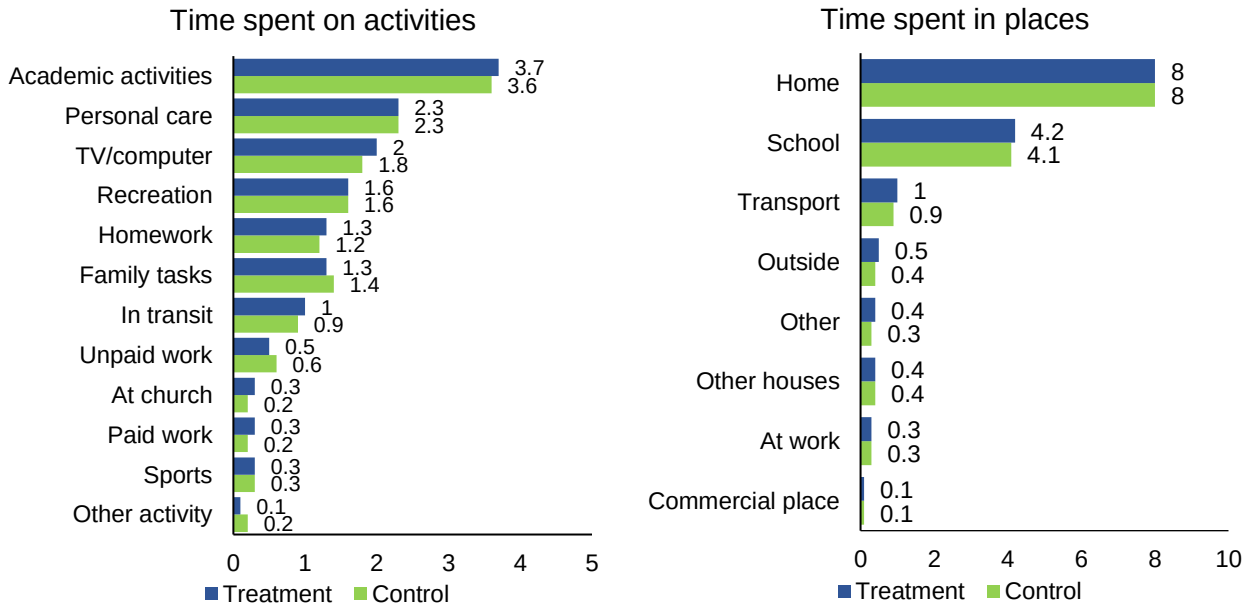
IV. STUDENTS' USE OF TIME, PERCEPTIONS OF TEACHING PRACTICES, CLASSROOM CLIMATE, AND SAFETY IN THE COMMUNITY

A. Baseline equivalence on students' use of time and perceptions

1. Students' use of time

We found no significant differences in third-cycle students' self-reported use of time on academic and nonacademic activities per day. Students in both the treatment and control groups reported that they spent four hours on average in class or involved in academic activities per day. They spent two hours on average tending to personal care,[10] watching television or involved in computer or recreation activities, and engaging in other activities, including recreation, homework, and family tasks. Consistent with these findings, students in both groups reported that they spent on average eight hours per day at home and four hours in school. We found no statistically significant differences in the time they spent on all activities or in the places where they spent time (Figure IV.1).

Figure IV.1. Third-cycle students' use of time in one day (hours)



Source: 2017 student survey.

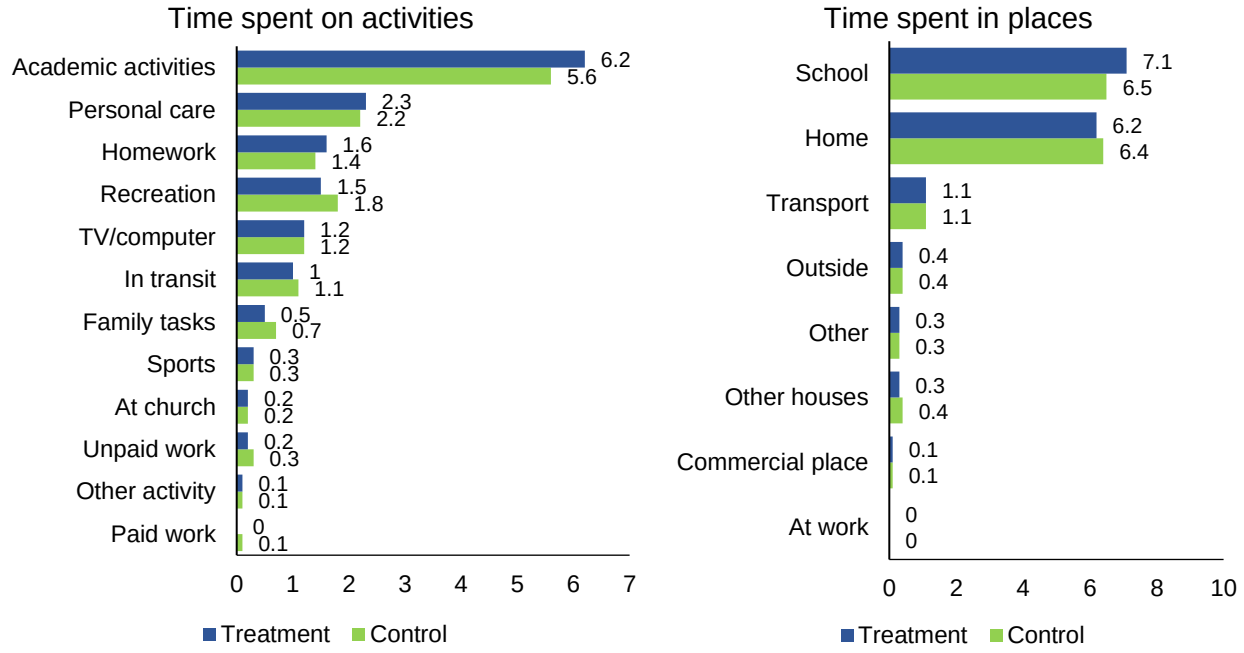
Note: The bars present group means that are regression-adjusted for the stratification used in the random assignment design.

In the treatment and control groups, students in secondary school spent a similar amount of time on academic and other activities as well as in school and other places. Secondary students in the treatment and control groups reported that they spent 6 hours on average in class or involved in academic activities, followed by time spent on personal care (2 hours on average) and homework (1.5 on average). We found no statistically significant

[10] Personal care includes activities such as personal hygiene, eating, and preparing meals or snacks.

differences in the time they spent on various activities (Figure IV.2). Furthermore, we found no statistically significant differences in the places where students spent time. Students in both groups spent approximately 7 hours at school and 6 hours a day at home. This was unexpected because secondary schools offer 8 hours per day of school. We explored this and found that the average time spent in school is lower than the expected 8 hours due to student absences. The student survey asked students to report their time use on a particular day, in this case, the day before the interview was conducted. Hence, the students who had been absent from school the day before the interview reported spending no time in school which lowered the average time in school across all students. .

Figure IV.2. Secondary students' use of time in one day (hours)



Source: 2017 student survey.

Note: The bars present group means that are regression-adjusted for the stratification used in the random assignment design.

We found statistically significant differences between the treatment and control groups in the percentage of students in third-cycle schools who reported that they spent more than four hours per week working on science and social science tasks. There were no differences for students in secondary schools. More third-cycle students in the control group versus third-cycle students in the treatment group reported spending more than four hours working on science tasks—the difference of 7 percentage points was statistically significant (Table IV.1). We also found that 22 percent of third-cycle students in the control group reported that they spent more than four hours per week on social science tasks in comparison with 16 percent of third-cycle students in the treatment group—this difference also was statistically significant. In contrast, for students in secondary, we found no statistically significant differences in the percentage of students who reported that they spent more than four hours per week on any subject area.

Table IV.1. Percentage of students who reported spending more than four hours per week working on the following subject areas

	Treatment group (A)	Control group (B)	Difference (A-B)	*p*-value
Third-cycle schools				
Science, health, and environment	17.3	24.0	-6.7*	0.039
Mathematics	17.0	19.4	-2.4	0.464
English	16.3	17.0	-0.8	0.776
Social science or civics	15.9	22.2	-6.4*	0.018
Language or literature	14.7	16.7	-2.0	0.531
Number of students	**539**	**617**		
Secondary schools				
Science, health, and environment	29.5	26.8	2.8	0.473
Mathematics	29.3	28.1	1.2	0.711
Social science or civics	28.4	25.6	2.9	0.448
Language or literature	24.6	22.4	2.3	0.499
English	19.8	17.2	2.7	0.408
Number of students	**340**	**400**		

Source: 2017 student survey.

Note: Column A and Column B present group means that are regression-adjusted for the stratification used in the random assignment design. Column A–B presents differences in the regression-adjusted group means between the treatment and control groups. The table includes p-values from tests of differences between group means.

* Difference in group means is statistically significant at the .05 level.

2. Perception of teachers' support and engagement in school

As discussed in Chapter I, we used subscales from the Classroom Life Scale (CLS) questionnaire to measure students' perceptions of teacher academic support and behavioral and emotional engagement in school, using items and the methodology developed by Van Ryzin, Gravely, and Roseth (2009). A detailed discussion of the constructed indicators appears in Appendix A. For the construction of the teacher academic support subscale,[11] we used the answers to four questions about academic support provided by the teachers, such as *my teachers want me to do my best in schoolwork*, with a 5-point Likert-type scale with 1 indicating never and 5 indicating always. We constructed behavioral and emotional engagement subscales[12] by using 20 self-reported items that assess the level of student engagement in classroom activities. Scores ranged between 1 (Very false) and 5 (Very true).

We found no statistically significant differences between the treatment and control groups in students' perceptions of teacher academic support or students' self-reported

[11] The teacher academic support subscale measures perceptions of support from teachers in relation to academic support.

[12] These subscales assess students' level of engagement in classroom activities related to behavioral engagement such as effort and attention and emotional engagement attitudes such as involvement and enjoyment.

emotional and behavioral school engagement. On average, students in third-cycle and secondary schools in both study groups reported that they received academic support from teachers most of the time. Teachers' academic support was reflected in students' feeling that their teachers were invested in their learning, wanted them to perform well, and helped them learn. Similarly, third-cycle and secondary school students in the treatment and control groups reported similar levels of school engagement. The average scores on the behavioral and engagement subscales of the CLS suggested that students were interested in school, participated in class, and made efforts to perform well on academic assignments. Table IV.2.

Table IV.2. Students' perception of teachers' support and engagement in school

	Treatment group (A)	Control group (B)	Difference (A-B)	*p*-value
Third-cycle schools				
Teacher academic support subscale	4.6	4.7	-0.1	0.080
Behavioral engagement subscale	4.1	4.0	0.1	0.193
Emotional engagement subscale	4.5	4.5	0.0	0.552
Number of students	**539**	**617**		
Secondary schools				
Teacher academic support subscale	4.5	4.6	-0.1	0.255
Behavioral engagement subscale	4.0	4.1	-0.1	0.266
Emotional engagement subscale	4.4	4.4	-0.0	0.474
Number of students	**340**	**400**		

Source: 2017 student survey.

Notes: Column A and Column B present group means that are regression-adjusted for the stratification used in the random assignment design. Column A–B presents differences in the regression-adjusted group means between the treatment and control groups. The table includes p-values from tests of differences between group means.

* Difference in group means is statistically significant at the .05 level.

3. Perception of classroom climate and teaching practices

With one exception, we found no statistically significant differences between the treatment and control groups in the percentage of students' perceptions about classroom climate and gender equity. Students reported how they felt in the classroom by agreeing or disagreeing with statements such as *I feel comfortable sharing my own perspectives and experiences in class* or *I feel marked for being a man or woman*. In third-cycle schools, we found only one statistically significant difference between the treatment and control groups in students' perceived academic competence. More students in the control group than in the treatment group agreed or completely agreed with the statement about feeling to work harder than other students to be perceived as a good student. The majority of students in both groups felt comfortable in sharing their perspectives in class (94 percent). However, 39 percent of students did not like to participate in class discussions. In addition, about 10 percent of students (male and female) agreed or completely agreed that they had heard teachers discriminate against students based on their gender. In secondary schools, we found similar results; the majority of students felt

comfortable in sharing their perspectives in class. However, 40 percent did not like to participate in class discussions, and 20 percent felt marked for their gender (male or female). We found no statistically significant differences between groups. Table IV.3.

Table IV.3. Students' perception of classroom climate and gender equity

	Treatment group (A)	Control group (B)	Difference (A-B)	*p*-value
Percentage of students who agree/completely agree with the following statements:				
Third-cycle schools				
I feel comfortable in sharing my own perspectives and experiences in class	95.9	93.0	2.9	0.074
I feel marked in class for being a male or female	22.8	26.2	-3.4	0.138
I feel that I have to work harder than other students to be perceived as a good student	70.0	75.6	-5.7*	0.046
In class, I've heard teachers discriminate against males or females	10.2	11.4	-1.2	0.548
I do not like to participate in class discussions	36.5	40.5	-4.0	0.191
Number of students	**539**	**617**		
Secondary schools				
I feel comfortable in sharing my own perspectives and experiences in class	93.0	95.4	-2.4	0.220
I felt marked in class for being a male or female	19.9	19.7	0.2	0.945
I feel that I have to work harder than other students to be perceived as a good student	64.3	58.7	5.6	0.259
In class, I've heard teachers discriminate against males or females	11.8	11.5	0.3	0.916
I do not like to participate in class discussions	40.6	37.6	3.1	0.418
Number of students	**340**	**400**		

Source: 2017 student survey.

Note: Column A and Column B present group means that are regression-adjusted for the stratification used in the random assignment design. Column A–B presents differences in the regression-adjusted group means between the treatment and control groups. The table includes p-values from tests of differences between group means.

* Difference in group means is statistically significant at the .05 level.

With a few exceptions, we found no significant differences in the percentage of students reporting that most of their teachers used positive teaching practices. Using a 4-point scale with 1 indicating a few teachers and 4 indicating all teachers, students reported on the proportion of teachers who performed 10 teaching practices. We constructed an indicator measuring the percentage of students reporting that most of their teachers used these teaching practices, all of which aimed to encourage or engage students in the classroom. In third-cycle schools, more control group students than treatment group students reported that most of their teachers shared personal stories or experiences in class; the difference of 5 percentage points was statistically significant. In secondary schools, more control group students (91 percent) than treatment group students (87 percent) reported that most of their teachers helped them learn how to make positive changes in society; the difference was statistically significant. In addition, more control group

students than treatment group students reported that most of their teachers valued the strengths and difficulties of the students in the classroom; the difference of 7 percentage points was statistically significant. Table IV.4

Table IV.4. Students' perception of teaching practices

	Treatment group (A)	Control group (B)	Difference (A-B)	*p*-value
Percentage of students who reported that most of their teachers:				
Third-cycle schools				
Encouraged students with different characteristics to work together	95.3	94.8	0.6	0.665
Helped students learn how to make positive changes in society	94.9	93.9	1.0	0.541
Encouraged students to contribute different points of view in class	93.2	93.7	-0.5	0.719
Motivated students to work harder than they thought they could	92.4	93.6	-1.2	0.428
Taught each student to be tolerant and respectful of different ideas or beliefs	91.1	93.5	-2.4	0.123
Valued the strengths and difficulties of the students in the classroom	89.4	90.5	-1.1	0.546
Took into account students' different abilities	89.4	88	1.5	0.444
Transformed controversial issues into meaningful discussions for students	80.4	79.3	1.1	0.672
Shared in class some personal stories or experiences	75.7	80.6	-4.9*	0.041
Spoke openly about issues of social inequality	72.7	72.3	0.4	0.884
Number of students	**539**	**617**		
Secondary schools				
Encouraged students with different characteristics to work together	92.8	91.8	1.1	0.641
Encouraged students to contribute different points of view in class	92.7	90.3	2.4	0.28
Taught each student to be tolerant and respectful of different ideas or beliefs	90.5	93.3	-2.8	0.182
Motivated students to work harder than they thought they could	90.3	92.2	-1.9	0.356
Helped students learn how to make positive changes in society	87.0	91.4	-4.4*	0.043
Took into account students' different abilities	85.7	87.5	-1.8	0.586
Valued the strengths and difficulties of the students in the classroom	81.8	88.6	-6.8*	0.041
Spoke openly about issues of social inequality	75.3	77.1	-1.9	0.604
Shared in class some personal stories or experiences	74.3	76.8	-2.5	0.473
Transformed controversial issues into meaningful discussions for students	73.4	73.7	-0.3	0.941
Number of students	**340**	**400**		

Source: 2017 student survey.

Note: Column A and Column B present group means that are regression-adjusted for the stratification used in the random assignment design. Column A–B presents differences in the regression-adjusted group means between the treatment and control groups. The table includes p-values from tests of differences between group means.

* Difference in group means is statistically significant at the .05 level.

With one exception, we found no significant differences between the treatment and control groups in the percentage of students reporting that their teachers prioritized certain tasks. Using a 5-point scale, students reported how important it was for teachers that students performed activities such as *memorize information, work in groups, or share opinions*. Unlike the previous questions, this question was not tied to guidelines for constructing an indicator. Instead, we constructed a binary indicator to demonstrate whether students reported the activity as important or very important (a value of 4 or 5) to teachers. The results in table IV.5 show the percentage of students who perceived that the activity was important or very important to teachers. In third-cycle schools, more than 90 percent of students reported that understanding assigned readings, studying and tending to homework, understanding the main ideas, and maintaining cordial relations with classmates were important to teachers. However, we found only one statistically significant difference between groups. More third-cycle students in the control group versus those in the treatment group reported that copying information from the blackboard was important to teachers; the difference of 5 percentage points was statistically significant. In secondary schools, most of the students reported that understanding main ideas, analyzing and interpreting ideas beyond what was read, maintaining cordial relations with classmates, and understanding the readings were important to the teacher. We found no statistically significant differences between study groups.

Table IV.5. Perception of teachers' support and engagement in school

	Treatment group (A)	Control group (B)	Difference (A-B)	*p*-value
Percentage of students who reported that their teachers prioritized the following activities:				
Third-cycle schools				
Understand the readings	93.7	93.1	0.6	0.736
Study and do homework	92.5	92.6	-0.1	0.970
Understand the main ideas	91.9	94.5	-2.6	0.056
Maintain cordial relations with classmates	90.7	91.3	-0.6	0.712
Prepare to study for a technical or professional career	89.9	87.8	2.2	0.303
Memorize information	89.3	89.1	0.2	0.937
Work in groups	88.4	88.9	-0.6	0.816
Share opinions or own ideas	88.1	88.1	0.0	0.988
Copy information from the blackboard	88.0	92.5	-4.5*	0.012
Analyze and interpret ideas beyond what is read	87.5	89.7	-2.2	0.193
Participate in sports activities	81.6	80.7	1.0	0.714
Participate in artistic activities	76.2	75.1	1.1	0.716
Use computers or technology for school work	67.9	63.4	4.5	0.259
Number of students	**539**	**617**		
Secondary schools				
Understand the main ideas	94.9	94.4	0.5	0.755
Analyze and interpret ideas beyond what is read	93.3	91	2.3	0.294
Maintain cordial relations with classmates	92.8	93.9	-1.1	0.522
Understand the readings	92.3	90.8	1.6	0.416

TABLE IV.5 *(Continued)*

	Treatment group (A)	Control group (B)	Difference (A-B)	*p*-value
Prepare to study for a technical or professional career	91.7	92	-0.3	0.903
Share opinions or own ideas	89.5	91	-1.5	0.455
Study and do homework	87.5	89.7	-2.2	0.489
Work in groups	85.5	89.8	-4.3	0.05
Use computers or technology for school work	81.7	82.3	-0.6	0.85
Copy information from the blackboard	80.6	86.7	-6.1	0.079
Memorize information	80.4	81.3	-0.9	0.774
Participate in artistic activities	67.2	68.4	-1.2	0.768
Participate in sports activities	64.6	68.9	-4.3	0.299
Number of students	**340**	**400**		

Source: 2017 student survey.

Note: Column A and Column B present group means that are regression-adjusted for the stratification used in the random assignment design. Column A–B presents differences in the regression-adjusted group means between the treatment and control groups. The table includes p-values from tests of differences between group means.

Adopted from High School Survey of Student Engagement (HSSSE), Center for Evaluation and Education Policy, Indiana University.

* Difference in group means is statistically significant at the .05 level.

4. Perception of community safety

We found no statistically significant differences between the treatment and control groups in third-cycle students' perceptions of safety in their community. In contrast, fewer treatment group students in secondary schools reported that they felt safe in their communities compared to control group students. Among the treatment and control group students enrolled in third-cycle schools, approximately 88 percent of students reported that they felt safe at school, 59 percent reported that they felt safe going to school, and 74 percent reported that they felt safe in their community. In secondary schools, 86 percent of treatment group students and 90 percent of control group students reported that they felt safe at school, with no statistically significant difference. But fewer students, especially in the treatment group, felt safe going to school— 45 percent of treatment group students reported that they felt safe going to school versus 55 percent of control group students; the difference of 10 percentage points was statistically significant. We also found that 65 percent of students in the treatment group reported that they felt safe in their community compared with 74 percent in the control group; the difference of 9 percentage points was not statistically significant at the 5 percent level but is close (Table IV.6). Students and principals also reported the frequency of problems such as theft of belongings and fights occurring in school. We found no statistically significant differences in these frequencies between the treatment and control groups. The detailed results appear in Appendix C.

Table IV.6. Percentage of students who reported that they felt safe in the community

	Treatment group (A)	Control group (B)	Difference (A-B)	*p*-value
Third-cycle schools				
Felt safe at school	87.8	88.2	-0.4	0.853
Felt safe going to school	61.7	56.3	5.4	0.120
Felt safe in their community	74.6	74.3	0.3	0.916
Number of students	**539**	**617**		
Secondary schools				
Felt safe at school	86.2	90.3	-4.1	0.124
Felt safe going to school	44.8	54.9	-10.1*	0.022
Felt safe in their community	64.9	73.6	-8.7	0.050
Number of students	**340**	**400**		

Source: 2017 student survey.

Note: Column A and Column B present group means that are regression-adjusted for the stratification used in the random assignment design. Column A–B presents differences in the regression-adjusted group means between the treatment and control groups. The table includes p-values from tests of differences between group means.

* Difference in group means is statistically significant at the .05 level.

B. Baseline equivalence by gender subgroup

One cross-cutting component of MCC's interventions is related to gender. In particular, FOMILENIO II's implementation of the SI-EITP model includes teacher training in gender equity. Therefore, FOMILENIO II and MCC are interested in understanding how male and female students use their time, as well as their perceptions on support from teachers, gender attitudes and school safety, before implementation of the SI-EITP. To this end, we examined by gender subgroup the indicators created from the student survey and tested separately the equivalence of the treatment and control groups among males and females. We also compared the group differences found among males with the group differences found among females.

Overall, the baseline equivalence results by gender subgroup were consistent with what we found in the full sample for third-cycle and secondary schools. In the analysis by gender, students' use of time was similar between study groups as were other student outcomes such as student engagement, perceived support from teachers, gender attitudes, and school safety. We found fewer statistically significant differences between the treatment and control groups among males compared to differences between study groups among females. For example, in third-cycle schools, we found a statistically significant difference in commuting time between females in the treatment and control groups, but no statistically significant difference for males. However, we do not think there is a consistent pattern in the differences that we found. We therefore conclude that the results for both females and males tended to be similar to those of the full sample for most outcomes. In Appendix C, we present the tables with separate baseline equivalence analyses for males and females.

This page has been left blank for double-sided copying.

V. EDUCATION INDICATORS FOR THIRD-CYCLE AND SECONDARY STUDENTS ENROLLED IN SCHOOL YEAR 2017 (BASE YEAR)

As described in Chapter I, we also used administrative data provided by MINED to assess baseline equivalence on education indicators such as dropout, repetition, and grade progression for third-cycle and secondary schools. The education indicators for third-cycle grades were created using two consecutive school years. For the analysis in secondary grades, we used data at the beginning and the end of the year. Table B.1 in Appendix B presents the description and methodological notes of the education indicators constructed.

In this chapter, we present our findings on the equivalence of education indicators using data for school years 2017 and 2018. In Appendix D, we present the baseline equivalence results for student indicators using data for school years 2016 and 2017 provided by MINED.

A. Baseline equivalence in third-cycle education indicators in base year 2017

In this section, we first describe the sample used to construct the education indicators for third-cycle schools, then discuss differences between treatment and control groups in students' characteristics and education indicators. Table V.1 shows the number of students enrolled in third-cycle schools in our study sample in the 2017 school year. The analysis sample excludes 40 third-cycle schools that did not provide any data for the 2018 school year. Hence, the analysis sample includes student data from 341 third-cycle schools that provided data for both 2017 and 2018 school years (out of the 381 schools in our evaluation sample).

Table V.1. Number of students enrolled in third-cycle grades at the beginning of the 2017 school year by study group

	Treatment group	Control group
Grade 7	5,395	5,976
Grade 8	4,984	5,445
Grade 9	4,637	5,131
Total number of students	15,016	16,552
Number of schools	158	183
Number of systems	45	53

Source: MINED enrollment registry for 2017 (SAE system).

Note: Two integrated systems in the control group are excluded due to having missing data: one did not provide data in 2018 and the other does not have third-cycle schools.

In third-cycle schools, the percentage of students who were overage was similar between study groups for grades 8 and 9 but not grade 7. We found a significant difference in the percentage of students who were overage[13] in 7th grade. The difference of two percentage points is statistically significant. In both groups, the percentage of students who were overage in

[13] Overage students are older than the official school-age range for the educational level they are enrolled in.

8th and 9th grade is similar. Approximately 14 percent of the students were overage in grade 8 and 12 percent were overage in grade 9 (Table V.2).

In third-cycle schools, the study groups had similar percentages of females in grades 7 and 9 but not in grade 8. As shown in Table V.2, the percentage of female students in the study groups was approximately 47 percent in grades 7 and 9. However, in grade 8, we found a statistically significant difference of 3 percentage points— in favor of the treatment group—in the percentage of female students.

Table V.2. Students' characteristics in base year 2017 for third-cycle schools

	Treatment group (A)	Control group (B)	Difference (A-B)	*p*-value
Percentage of overage students				
Grade 7	13.8	16.2	2.4*	0.04
Grade 8	14.1	14.5	0.4	0.71
Grade 9	11.6	13.0	1.4	0.14
Percentage of female students				
Grade 7	45.3	46.3	1.0	0.30
Grade 8	47.6	45.0	-2.6*	0.02
Grade 9	48.0	47.4	-0.6	0.66

Source: MINED (SAE system) for school year 2017.

Note: Column A and Column B present group means that are regression-adjusted for the stratification used in the random assignment design. Column A-B presents differences in the regression-adjusted group means between the treatment and control groups. The table includes *p*-values from tests of differences between group means are also included in the table.

* Difference in group means is statistically significant at the .05 level.

In third-cycle schools, educational indicators such as progression to the next grade, dropout across years, and grade repetition were similar in both groups. In Table V.3, we present education indicators from 2017 to 2018 for students enrolled in third-cycle at the beginning of the 2017 school year. Seventy-two percent of students enrolled in grade 7 at the beginning of 2017 progressed to grade 8 in the same school in the next school year, 22 percent dropped out of the same school in the next school year, and 5 percent repeated grade 7 in the same school in the next school year. Of the students enrolled in grade 8 in the 2017 school year, 75 percent progressed to grade 9 in the same school in 2018, 21 percent dropped out from the school, and 4 percent repeated 8th grade in the same school in the next school year. We found no statistically significant differences between treatment and control groups on any of these indicators.

Transition from third-cycle to secondary was similar in both groups. The transition between third-cycle and secondary occurs from 9th grade to 10th grade. We found no statistically significant differences between groups in the percentage of students enrolled in grade 9 in 2017 who were enrolled in grade 10 in any secondary school in the next school year, 2018.

Table V.3. Education outcomes from 2017 school year to 2018, by study group

	Treatment group (A)	Control group (B)	Difference (A-B)	*p*-value
Dropout rate (across years in the same school)				
Grade 7	23.3	21.5	1.8	0.462
Grade 8	21.7	20.0	1.7	0.611
Progressed to the next grade (in the same school)				
Grade 7	71.8	72.7	-0.9	0.702
Grade 8	74.5	76.2	-1.7	0.616
Repeated grade (in the same school)				
Grade 7	4.9	5.8	-0.8	0.228
Grade 8	3.8	3.8	0.0	0.932
Grade 9	1.4	1.7	-0.3	0.392
Transition from 9th to 10th grade				
Percentage of students enrolled in 2017 in grade 9 who were enrolled in grade 10 in any secondary school in 2018[14]	69.8	73.1	-3.2	0.318

Source: MINED (SAE system) for school year 2017.

Note: Column A and Column B present group means that are regression-adjusted for the stratification used in the random assignment design. Column A-B presents differences in the regression-adjusted group means between the treatment and control groups. The table includes *p*-values from tests of differences between group means are also included in the table.

* Difference in group means is statistically significant at the .05 level.

B. Baseline equivalence in secondary educational outcomes in base year 2017

Table V.4 presents student enrollment disaggregated by type of baccalaureate for the schools in the sample for which MINED made the 2017 school year data available.

[14] This indicator includes all students enrolled in grade 10 regardless of whether the secondary school was in the same Integrated System or not.

Table V.4. Number of students enrolled at the beginning of school year 2017, by study group

	Treatment group	Control group
General		
Grade 10	1,867	2,005
Grade 11	1,579	1,767
Technical		
Grade 10	1,341	2,359
Grade 11	1,128	1,921
Grade 12	1,181	1,649
Total number of students	7,096	9,701
Number of secondary schools	39	51
Number of integrated systems	33	40

Source: MINED enrollment registry for 2017 (SIRAI system).

In secondary schools, the likelihood of being overage was similar in both groups for all grades. In grade 10, approximately 13 percent of students in general programs and 11 percent in technical programs were overage. In grade 11, 12 percent of students in general programs and 8 percent in technical programs were overage. The differences between groups are not statistically significant (Table V.5).

In secondary schools, study groups had similar percentages of females with one exception. In general programs, the percentage of female students in the study groups was approximately 47 percent in grades 10 and 11. The differences between groups are not statistically significant. In technical programs, study groups had similar percentage of females in grade 10 (47 percent) and grade 12 (51 percent). In grade 11, we found a statistically significant difference of 5 percentage points in the percentage of female students in favor of the treatment group (Table V.5).

Table V.5. Students' characteristics by study group and grade in 2017

	Treatment group (A)	Control group (B)	Difference (A-B)	p-value
Percentage of overage students				
General				
Grade 10	12.5	12.7	-0.2	0.898
Grade 11	11.4	11.6	-0.3	0.847
Technical				
Grade 10	9.9	10.8	-0.9	0.673
Grade 11	8.0	7.7	0.2	0.857
Grade 12	8.8	10.3	-1.5	0.289
Percentage of female students				
General				
Grade 10	45.4	46.9	-1.4	0.457

TABLE V.5 *(Continued)*

	Treatment group (A)	Control group (B)	Difference (A-B)	p-value
Grade 11	48.1	47.5	0.6	0.767
Technical				
Grade 10	45.5	48.2	-2.6	0.313
Grade 11	52.5	47.6	4.9*	0.007
Grade 12	50.4	51.7	-1.3	0.544

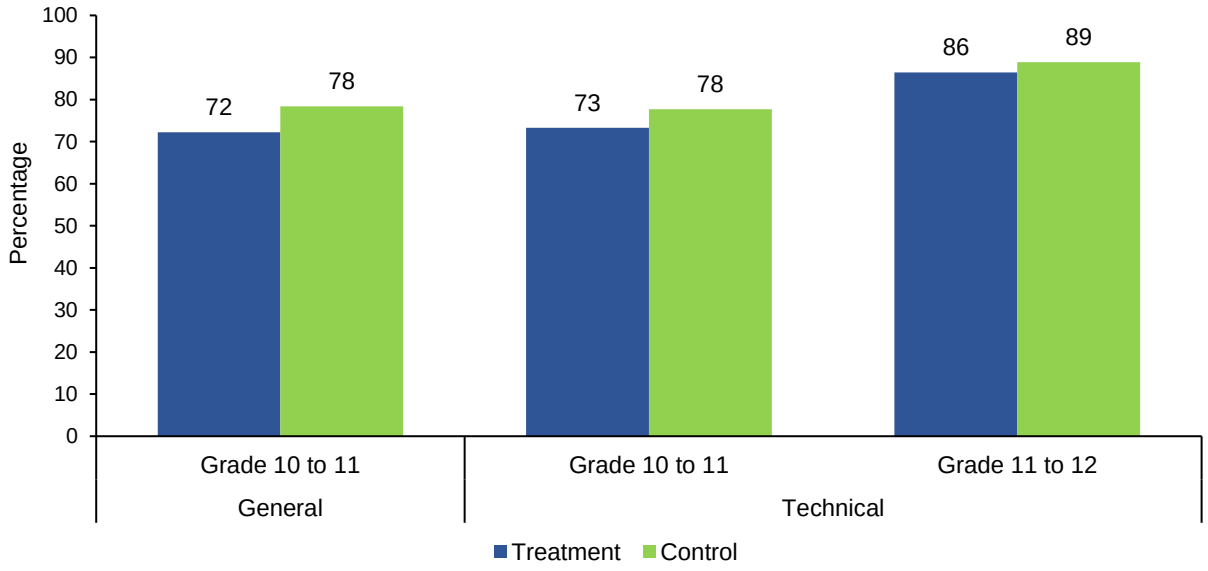
Source: MINED (SIRAI system) for school year 2017.

Note: Column A and Column B present group means that are regression-adjusted for the stratification used in the random assignment design. Column A-B presents differences in the regression-adjusted group means between the treatment and control groups. The table includes *p*-values from tests of differences between group means are also included in the table.

* Difference in group means is statistically significant at the .05 level.

In secondary schools, student progression to the next grade in the same school was similar in both groups. A student progresses to the next grade in the same school when a student enrolled in a certain grade in 2017 is enrolled in the next grade in 2018 in the same school. Among students enrolled in grade 10 in technical and general programs in 2017, 72 percent of students in the treatment group and 78 percent in the control group progressed to grade 11 in the same school in the 2018 school year. For students enrolled in grade 11 in technical programs, 86 percent of students in the treatment group and 89 percent in the control group progressed to grade 12. None of the differences between study groups is statistically significant (Figure V.1).

Figure V.1. Progression to the next grade (in the same school), by study group in 2017



Source: MINED (SIRAI system) for beginning and end of school year 2017.

Note: The bars present group means that are regression-adjusted for the stratification used in the random assignment design.

In secondary grades, dropout rates were similar across groups except in grade 10 in the general program. As shown in Table V.6, among students enrolled in 2017 in grade 10 in the general program, 24 percent of students in the treatment group and 17 percent in the control group dropped out from that school in the next school year, 2018. This difference was statistically significant. We found no significant differences among study groups in dropout rates across years among students enrolled in grade 10 in technical programs. For students enrolled in grade 11 in technical programs, 11 percent of students in the treatment group and 10 percent in the control group dropped out from the school in the next school year.

Grade repetition rates were similar in both study groups. In grade 10, the repetition rate was 4 percent in general programs and approximately 6 percent among students in technical programs. In grade 11, the repetition rate was 1 percent in the treatment group and 2 percent in the control in general programs. The repetition rate among students enrolled in technical programs was 3 percent in the treatment group and 1 percent in the control group. These differences are not statistically significant at the 5 percent level.

We found no statistically significant differences between groups in the percentage of students who passed their grade at the end of the school year (regardless of enrollment in next grade) in secondary schools. Using beginning and end of 2017 school year data, we estimated the percentage of students who passed or failed the grade by type of program (general or technical baccalaureate). Approximately 80 percent of students enrolled in grade 10 in 2017 passed the grade in both types of programs. Among students enrolled in grade 11, 90 percent passed their grade in general and technical programs. Nearly all students enrolled in grade 12 in technical programs passed that grade (Table V.6).

Table V.6. Education outcomes in 2017, by study group

	Treatment group (A)	Control group (B)	Difference (A-B)	*p*-value
Dropout rate (across years) in the same school at the end of the school year				
General				
Grade 10	24.0	17.1	6.9*	0.020
Technical				
Grade 10	21.1	17.0	4.1	0.156
Grade 11	11.0	9.7	1.3	0.635
Repeated grade (in the same school)				
General				
Grade 10	3.7	4.5	-0.7	0.528
Grade 11	0.6	1.5	-0.9	0.082
Technical				
Grade 10	5.6	5.3	0.3	0.777
Grade 11	2.5	1.4	1.1	0.069
Grade 12	0.7	0.5	0.2	0.665
Pass rate (at the end of the school year)				
General				

TABLE V.6 *(Continued)*

	Treatment group (A)	Control group (B)	Difference (A-B)	*p*-value
Grade 10	77.7	83.5	-5.8*	0.048
Grade 11	89.3	91.1	-1.8	0.391
Technical				
Grade 10	78.7	81.4	-2.7	0.270
Grade 11	88.7	90.9	-2.1	0.382
Grade 12	96.1	96.8	-0.7	0.413

Source: MINED (SIRAI system) for beginning and end of school year 2017.

Note: Column A and Column B present group means that are regression-adjusted for the stratification used in the random assignment design. Column A-B presents differences in the regression-adjusted group means between the treatment and control groups. The table includes *p*-values from tests of differences between group means are also included in the table.

* Difference in group means is statistically significant at the .05 level.

Baseline equivalence in PAES scores

MINED also provided student-level data of the scores obtained by students enrolled in 11th grade in the Learning and Skills Test for Graduates of Secondary Education (PAES) for the school years 2016 and 2017. PAES is scored on a 10-point scale to assess students' knowledge in four areas: mathematics, social studies, language, and science. Table V.7 shows the scores for the school year 2017. In Appendix D, we present the PAES scores for school year 2016.

We found no significant differences between study groups in the scores of the PAES standardized test in any of the subject areas assessed. In both groups, the global score was around 5, with the highest score obtained in language (5.8), followed by social science (5.5), science (5.4), and mathematics was considerably lower (4.5). The scores from students in our study sample are similar to the scores at the national level. As a reference, at the national level, the global score in 2017 was 5.3, language score was 6.0, social science was 5.8, science was 5.5, and mathematics was also considerably lower 4.8.

Table V.7. PAES scores by subject area and study group in 2017

	Treatment group (A)	Control group (B)	Difference (A-B)	*p*-value
Global	5.1	5.1	0.0	0.986
Language	5.9	5.7	0.2	0.105
Social science	5.5	5.5	-0.0	0.995
Science	5.4	5.3	0.1	0.439
Mathematics	4.5	4.8	-0.3	0.144
Number of students	**2,489**	**3,381**		

Source: MINED for school year 2017.

Note: Column A and Column B present group means that are regression-adjusted for the stratification used in the random assignment design. Column A-B presents differences in the regression-adjusted group means between the treatment and control groups. The table includes *p*-values from tests of differences between group means are also included in the table.

* Difference in group means is statistically significant at the .05 level.

VI. CONCLUSIONS

A. Summary and conclusions

In this report, we document baseline characteristics and verify balance in study group observable characteristics. We use survey data collected in schools in the school year 2017 and administrative data from MINED for school years 2016 to 2018.

The main conclusion is that both groups had similar characteristics for schools, teachers, and students. Although we found a few significant differences, there is no evidence that one group had better characteristics than the other in general.

Both groups had similar facilities or services before the intervention started. In 2017, before the intervention started, third-cycle and secondary schools in both groups had basic facilities such as piped water supply and working toilets. However, both groups also lacked some basic academic resources, such as computer room for students, library or resource room, and science lab. The only difference we found was in access to running water in secondary schools, with treatment schools more likely to have piped water compared to control schools.

Schools in the treatment group had started implementing some key components of SI-EITP model. Although we found no statistically significant differences across study groups in principal-reported participation in the SI-EITP model, some components of the intervention had been implemented in the treatment group. For example, for extension of the school day, which is part of the SI-EITP model, 59 percent of treatment and 39 percent of control principals reported that the schools were already offering activities in extended time in the 2017 school year. In addition, third-cycle schools from the treatment group were more likely than the control group to report having participated in the pedagogical proposal. Hence, even before FOMILENIO II began implementing the intervention, components of SI-EITP, such as working together in a pedagogical proposal and extending some hours in third-cycle schools, had been implemented by MINED. This is not surprising given that the random assignment was done in 2016 and MINED started implementing a few components of the model in the intervention group. However, we expect that FOMILENIO II will implement all the components of the SI-EITP model stating in 2018.

We do not find evidence that the teachers from third-cycle and secondary schools were systematically different across study groups. Teachers in both groups had similar age, gender, experience, and education. Regarding training, teacher-reported participation in training focused on subject areas was similar in both groups. Moreover, we found no significant differences in the reported participation in training focused on gender equality, ICTs, class management, soft skills, differentiated instruction, or English language instruction. However, third-cycle teachers in the treatment group were more likely to have had training focused on active learning approaches and inclusive classrooms than teachers in the control group. Teachers from both groups reported performing teaching practices such as presenting information on the blackboard, solving exercises in pairs or small groups, and promoting group discussion with similar frequency.

Principals' participation in professional development activities was similar across study groups in third cycle but not in secondary. Although we found a few significant

differences between treatment and control groups in the reported participation of principals in training, we do not interpret them as evidence of systematic differences between principals in the treatment and control groups in third cycle. More principals from third-cycle treatment schools reported participating in training focused on gender equality than principals in the control group. However, principals from secondary schools in the control group were more likely to have had training in administrative management, technological areas, leadership, differentiated instruction, and staff management compared to principals in the treatment group. FOMILENIO II's implementation of the SI-EITP model includes intensive training for principals and school committees, we therefore expect the treatment group to receive more training than the control group in the coming school years. Our final models, will not account for these initial differences in principals' trainings because our objective in the follow-up analysis is to assess whether the principals in the treatment group received the planned training.

Teachers' use of class time at baseline was similar between study groups. Before the intervention started, teachers in the two study groups distributed class time in similar ways, spending around three-quarters of class time on academic activities and about 15 percent of class time on management activities. Teachers in both groups spent approximately 10 percent of class time off-task, which is not best practice. Teachers in both groups were observed using learning material in similar ways, relying much of their time on the blackboard and lecturing without any material. Finally, we found no differences between groups on student engagement. The percentage of time that students were engaged when the teacher was performing an academic activity was similar between groups.

Treatment and control students in the treatment and control groups had similar time use, student engagement, perceived support from teachers, gender attitudes, and school safety. Most of the student outcomes discussed are similar between groups. However, we found few statistical differences between perceptions of classroom climate and teaching practices. Finding a few differences is not surprising given that we tested a large number of variables and would expect to see some significant differences just by chance. For example, with a 5 percent significance for every 100 tests conducted, we expect to see five statistically significant differences by chance. Therefore, we do not perceive a consistent pattern in the differences found. Our final estimation models will not control for the few initial differences found in students' perceptions because there is no consistent evidence that students have differing perceptions between groups.

According to MINED's administrative data, schools in both study groups had similar education outcomes. In third-cycle schools, progression to the next grade, dropout across years, and grade repetition in the same school were similar in both groups. We also found that transition from third-cycle to secondary was similar in both groups. In secondary schools, treatment students enrolled in grade 10 in general programs were more likely to drop out from the same school across years than control students. Our final analysis will account for the initial difference in 10th grade dropout for general programs. However, for students enrolled in technical programs, the differences between study groups are not statistically significant. We also found no differences in the scores of the standardized test in any of the subject areas assessed.

The final analysis of teacher and student data will not account for baseline characteristics. The impact estimation of teacher and student outcomes will be done by comparing these outcome indicators in the treatment and control groups without accounting for baseline characteristics. We selected this estimation method taking into account the following. 1) At baseline, the treatment and control groups are similar in terms of school infrastructure and educational outcomes, teachers' characteristics, participation in professional development, use of class time, and students' use of time. 2) The intervention aims to improve infrastructure and training for principals and teachers in the treatment group; and our objective with the follow-up analysis is to assess whether the treatment group received the planned intervention. 3) The sample of teachers and students that will be interviewed at follow up is different than the baseline sample. 4) The survey instruments for teachers and students will be modified to better reflect current implementation plans.

The final analysis of the Stallings observations and educational outcomes from MINED's data will account for the baseline information to increase the statistical power. In this study, we shown that in general, the treatment and control groups were similar in terms of educational outcomes obtained from MINED data. Furthermore, the groups were similar in terms school, teacher, student and classrooms indicators obtained from surveys and classroom observation data. At follow-up, we will collect student level MINED's data on the same sample of schools we have at baseline. Therefore, we can follow the school's educational outcomes longitudinally. We will take advantage of having data on school outcomes at several points in time, to reduce the variance of the impact estimation. As discussed in the design report, (Campuzano et. al 2018), the impact estimation will use a regression analysis that will account for initial educational outcomes to reduce the variance of the impact estimate and obtain more precise estimates. Similarly, for the outcomes of the classroom observation using the Stallings method, we will take advantage of having baseline data for a sample of classrooms in the same schools. We will calculate the average baseline indicators at the school level and we will use them as covariates in the regression models in order to reduce the outcome variance and increase statistical power.

REFERENCES

Bruns, B., Costa, L; and N. Cunha. "Through the Looking Glass: Can Classroom Observation and Coaching Improve Teacher Performance in Brazil?" Policy Research working paper; no. WPS 8156; Impact Evaluation series. Washington, D.C.: World Bank Group, July 2017.

Bruns, B; Luque, J. "Great Teachers: How to Raise Student Learning in Latin America and the Caribbean." Washington, DC: World Bank. © World Bank. 2015. Available at https://openknowledge.worldbank.org/handle/10986/20488 License: CC BY 3.0 IGO."

Campuzano L., I. Padilla, and C. Fernández. "Design Report for the Impact and Performance Evaluation of the Education Quality Activity, El Salvador" Washington, DC: Mathematica Policy Research, December 2018.

Dirección Nacional de Educación Media (Tercer ciclo y Media) Ministerio de Educación. "Boletín de resultados Prueba de Aprendizaje y Aptitudes para Egresados de Educación Media (PAES)." Ministerio de Educación, 2017. Available at https://www.mined.gob.sv/paes/2017/boletines/Bolet%C3%ADn%20Informativo%20PAES%202017.pdf.

George, D., and P. Mallery. *SPSS for Windows Step by Step: A Simple Guide and Reference*, 11.0 update (4th ed.). Boston: Allyn & Bacon, 2003.

Joint USG-GOES technical team. "Partnership for Growth: El Salvador Constraint Analysis." 2011.

Ministerio de Educación, Gobierno de El Salvador. "Modelo de escuela inclusiva de tiempo pleno en El Salvador. Documento de Sistematización." Available at https://www.fomilenioii.gob.sv/modelo-de-escuela-inclusiva-de-tiempo-pleno-en-el-salvador--2.

Schuh Moore, A.M., J. DeStefano, and E. Adelman. "Using Opportunity to Learn and Early Grade Reading Fluency to Measure School Effectiveness in Ethiopia, Guatemala, Honduras, and Nepal." Washington, DC: EQUIP2, AED, and USAID, 2010.

Schochet, P.Z. "Statistical Power for School-Based RCTs with Binary Outcomes," *Journal of Research on Educational Effectiveness*, vol. 6, 2013, pp. 263–294.

Stallings, J., and S. Knight. "Using the Stallings Observation System to Investigate: Time on Task in Four Countries." Unpublished paper for the International Time on Task Project, World Bank, Washington, DC. 2003.

Tavakol, M., and R. Dennick. "Making Sense of Cronbach's Alpha." *International Journal of Medical Education*, vol. 2, 2011, pp. 53–55.

USAID. "Sistematización del proceso de elaboración del sistema integrado como estrategia de expansión del modelo de escuela inclusiva de tiempo pleno (EITP), para que sirva de guía al Ministerio de Educación de El Salvador (MINED), en la implementación de dicha estrategia. Informe final." Programa de fortalecimiento de la educación Básica EQUIP2, El Salvador Sistema Integrado de la Escuela Inclusiva de Tiempo Pleno. USAID, 2012.

Van Ryzin, M.J., A.A. Gravely, and C.J. Roseth. "Autonomy, Belongingness, and Engagement in School as Contributors to Adolescent Psychological Well-Being." *Journal of Youth and Adolescence*, vol. 38, no. 1, 2009, pp. 1–12.

APPENDIX A: OUTCOME DEFINITIONS

This page has been left blank for double-sided copying.

A. Key indicators for the impact evaluation of the SI-EITP

Table A.1 presents a summary of the outcome indicators constructed for the baseline analysis, the data sources used, and the domains to which they belong.

Table A.1. Descriptions and data sources of outcome indicators for the impact evaluation of the SI-EITP

Outcome indicator	Description	Data source	Domain
Short-term outcomes			
Schools share resources	Binary indicator identifying whether schools in the systems shared resources during the school year	Principal survey	Integrated System administration
Joint decision making across schools in the Integrated System	Binary indicator identifying whether schools in the systems made joint decisions regarding the pedagogical proposal during the school year	Principal survey	Integrated System administration
Joint decision making within schools	Binary indicator identifying whether teachers within schools made joint decisions regarding the pedagogical proposal during the school year	Principal and teacher surveys	School administration
Perceived parental and community involvement	Indicator identifying the perceived parental and community involvement	Principal and teacher surveys	Community involvement
Time spent on academic activities	Indicators that estimate share of total class time used on academic activities through a systematic observation of classroom interactions	Classroom observation (Stalling Observation)	Time on task
Time in school for students	Average number of hours that students spent at school in instructional/academic activities, sports, and recreational or vocational activities organized by the school and with teacher/adult supervision (based on student self-report on a subsample of 7th- and 9th-grade students)	Student survey	Time in school
Instructional practices	Indicators that estimate share of total class teachers' use of instructional time, use of material, core pedagogical practices, and share of class time students are engaged	Classroom observation (Stalling Observation)	Instructional practices
Alignment of teacher assignment with teacher training	Binary indicator identifying whether the teacher is teaching the subject for which he or she was trained	Teacher survey	School administration
Medium-term outcomes			
Enrollment in grade	The grade the student was enrolled in each year. Secondary schools have data for grades 10–12; third-cycle schools have data for grades 7–9.	Student records for third-cycle (SAE) and secondary (SIRAI)	Go to school
Dropped out (inter-year in the same school)	Binary indicator identifying whether a student who was enrolled in one school year was enrolled in the next school year in the same school. This indicator can be constructed using two consecutive years of MINED student-level data for grades 7–9 and 10–12.	SAE for grades 7–9 and SIRAI for grades 10–12	Leave school

TABLE A.1 *(Continued)*

Outcome indicator	Description	Data source	Domain
Repeated grade (in the same school)	Whether a student who was enrolled in a certain grade enrolled in the same grade the next year (in the same school). We construct this outcome from student-level grade enrollment data in two consecutive school years for grades 9–12.		Repetition
Passed grade	Whether a student passed the grade in which he or she was enrolled that year. Using the MINED student-level records, we can only construct this indicator with secondary schools' data for grades 10–12.		Progress in school (secondary outcome)
PAES test scores	Students' PAES grade 11 test scores for math, language, social studies, and science.	MINED records	Academic achievement
Transition from 9th to 10th grade (within or across schools inside or outside the Integrated System)	Binary outcome indicator from student-level enrollment data identifying whether students who were enrolled in 9th grade in any school in the system enrolled in 10th grade in the next year, regardless of whether the secondary school is in the same Integrated System or not. To construct the indicator, we will use the identification number of students enrolled in 9th grade in the system and track their enrollment in 10th grade in the following school year. In addition to the binary variable described above, we will construct another indicator for identifying students who enrolled in a secondary school belonging to the Integrated System.	School records from third-cycle schools; school records from secondary schools; visits to households SAE and SIRAI data from two consecutive school years	Progress in school

B. Description of indicators to measure teacher support, engagement in school, and perceived teaching practices

Table A.2 presents a summary of the outcome indicators constructed from the survey used to measure perceived academic support, engagement in school, perceived teaching practices, and perceived safety.

Table A.2. Description of indicators constructed from student survey

Indicator	Description	Construct	Data Source
Teacher academic support	Students answered 4 items about perceived teacher academic support such as *My teachers want me to do my best in schoolwork*. Students responded to each item using a 5-point scale from never (1) to always (5).	Item scores were averaged to obtain a subscale score of teacher academic support.	Adapted from the Classroom Life subscale of teacher academic support following method suggested by Van Ryzin et al. (2009)
Behavioral engagement	Students answered 10 items about effort and attention such as *I try hard to do well in schoo*l or *In class at school, I do just enough to get by*. Students responded using a 5-point scale from not at all true (1) to very true (5).	Behavioral item scores were added to obtain subscale scores, with negatively worded items being subtracted from positively worded items.	Adapted from the modified Engagement vs. Disaffection with Learning Scale as in Van Ryzin et al. (2009).

TABLE A.2 *(Continued)*

Indicator	Description	Construct	Data Source
Emotional engagement	Students answered 10 items about interest and enjoyment in school such as *When I'm in class at school, I feel good* and *When we work on something in class at school, I feel bored.* Students responded using a 5-point scale from not at all true (1) to very true (5).	Emotional item scores were added to obtain subscale scores, with negatively worded items being subtracted from positively worded items.	
Classroom climate and teaching practices	Students answered 10 items on how they feel in the classroom by agreeing or disagreeing with statements such as *I feel comfortable sharing my own perspectives and experiences in class* or *I feel marked for being a man or a woman*. Students responded to each item using a 4-point scale from completely disagree (1) to completely agree (4).	Binary variables were constructed for each item. The indicators show the percentage of students who agreed with the statements.	
Students' perception of teaching practices	Students answered 10 items about teachers who perform teaching practices aimed at encouraging or engaging them in the classroom, such as *Teacher encourage students with different characteristics to work together*. Students used a 4-point scale where 1 is a few teachers and 4 is all.	We constructed a binary indicator for each item. The indicators show the percentage of students who reported that most of their teachers perform the activities.	
Perception of prioritized teaching practices	Students answered 10 items about how important it was for teachers that students perform activities such as *memorize information, work in groups*, or *share opinions*. Students used a 5-point scale where 1 is not important and 5 is very important.	We constructed a binary indicator for each item. The indicator shows the percentage of students who reported it being important for teachers that students perform activities.	Adapted from High School Survey of Student Engagement (HSSSE), Center for Evaluation and Education Policy, Indiana University
Perception of community safety	Students answered 3 items about feeling safe in school, going to school, and in the community. Students used a 3-point scale where 1 is yes, 2 is more less and 3 is no.	We constructed a binary indicator for each item. The indicator shows the percentage of students who reported feeling safe.	

Internal consistency and reliability

Cronbach's alpha is a measure of internal consistency and reliability for multiple-item tests. It calculates the intercorrelation between test items; the higher the coefficient, the more the items measure a given concept in the same way (Tavakol and Dennick 2011). Cronbach's alpha scores range from 0 to 1. Scores close to 0 indicate that items are uncorrelated within the test and scores close to 1 show when items are perfectly correlated. The literature on Cronbach's alpha cites 0.6 to 0.7 as an acceptable range for establishing internal consistency within test items (George and Mallery 2003).

The internal consistency reliability of the three subscales, as measured by the Cronbach's alpha coefficient, was acceptable (within 0.6–0.7 range). We retained all items for each subscale.

Alpha coefficients did not improve substantially by removing any of the items. For the purpose of summarizing and presenting students' survey responses to the items from the three scales, we used the subscales as in Van Ryzin et al. (2009), although the alpha coefficients are somewhat lower than those reported for the original subscales.

Table A.3. Internal consistency and reliability (Cronbach's alpha)

	Value
Third-cycle schools	
Teacher academic support	0.6
Behavioral engagement subscale	0.7
Emotional engagement subscale	0.7
Number of students	**1,156**
Secondary schools	
Teacher academic support	0,6
Behavioral engagement subscale	0.7
Emotional engagement subscale	0.7
Number of students	**740**

APPENDIX B: SURVEY RESPONSE RATES AND ANALYTIC METHODS

This page has been left blank for double-sided copying.

A. Sample for data collection and response rates

In October 2017, DIGESTYC administered baseline principal surveys, teacher surveys, and student questionnaires, as well as baseline classroom and school observations made before the SI-EITP began implementation. In-school data collection will also include follow-up surveys to principals, teachers, and students, along with school and classroom observations, after the main components of the SI-EITP model are implemented (extension of the school day, teacher training, and secondary schools in each Integrated System).

The initial plan was to collect data in 273 schools: two third-cycle schools and one secondary school per Integrated System. However, DIGESTYC did not have the capacity to collect data from the entire target sample before the end of the school year; therefore, the target sample size was reduced to 190 schools—116 third-cycle and 74 secondary schools. For the reduced sample, one third-cycle school and one secondary school were randomly selected in each Integrated System.

Table B.1. Response rates in the base year (percent)

	Principal survey response rate	Classroom observation response rate	Teacher survey response rate	Student survey response rate
Third-cycle schools				
Treatment	100	98.8	100	100.0
Control	100	99.5	100	99.8
Secondary schools				
Treatment	100	100.0	100	99.4
Control	100	95.0	100	100.0

B. Analytic methods

Given the use of random assignment, the basic method to estimate differences between study groups in measurable characteristics of schools, teachers, classrooms, and students is to compare the mean outcomes of the treatment and control groups. We used regression models to estimate differences because the regression adjustment enabled us to account for the stratification used in the random assignment design.

We estimated differences between study groups using a regression specification that compares outcomes of students who attended a school that implemented the SI-EITP model (treatment group) with outcomes of students who attended schools that were not part of a potential integrated system (control group), controlling for idiosyncratic differences between the two groups. The regression model in the base year can be expressed as follows:

$$(1) \quad Y_{ipsd} = \alpha + \lambda T_s + \gamma z_d + \varepsilon_{ipsd}$$

In this equation, Y_{ipsd} is the characteristic or indicator of interest for student *i* in educational program or school *p* in Integrated System *s* in stratum *d* used in the random assignment design; T_s is an indicator equal to one if the student is in an educational program or school that is part of

an Integrated System *s* assigned to the treatment group (SI-EITP) and zero if the student is in a school that is part of an Integrated System assigned to the control group. The vector z_d includes indicator variables for each stratum *d* used in the random assignment design (department and the presence of a secondary school in the Integrated System). In addition, ε_{ipsd} is a random error term for student *i* in program or school *p* in Integrated System *s* in stratum *d*. The coefficient λ estimates the difference between the treatment and control group comparison. Finally, all of the estimates are weighted to account for differential assignment probability within stratum (department and the presence of a secondary school in the Integrated System). Because we conducted random assignment at the Integrated System level, we will adjust the standard errors for clustering at that level.

We also conducted an impact analysis by gender. In this instance, we ran the regression model for the sample of male and female students separately. These results are presented in Appendix D.

APPENDIX C: ADDITIONAL TABLES AND FIGURES

This page has been left blank for double-sided copying.

In this appendix, we present additional tables and figures from the analysis of principal, teacher, and student surveys.

A. Baseline equivalence of additional outcomes from teacher survey

We found no differences between study groups in the percentage of math and science teachers trained in these two subject areas; however, more language teachers from third-cycle schools were trained to teach language. Language teachers from the treatment group were more likely to be trained in teaching language compared with control language teachers. The difference of 21 percentage points is statistically significant. In secondary schools, we found no differences between study groups.

Table C.1. Teachers trained in the subject area they taught in the classroom observation (%)

	Treatment group (A)	Control group (B)	Difference (A-B)	*p*-value
Third-cycle schools				
Teachers trained in the subject area they taught when we observed them	60.0	48.7	11.3	0.054
Teachers trained in the subject area they taught when we observed them, by subject				
Mathematics	65.2	61.4	3.8	0.722
Science	53.5	44.3	9.2	0.365
Language	59.5	38.9	20.5*	0.045
Number of teachers	**144**	**167**		
Secondary schools				
Teachers trained in the subject area they taught when we observed them	81.7	74.7	7.0	0.261
Teachers trained in the subject area they taught when we observed them, by subject				
Mathematics	93.2	76.4	16.8	0.108
Science	84.0	67.5	16.5	0.119
Language	68.7	78.3	-9.6	0.420
Number of teachers	**99**	**114**		

Sources: 2017 teacher survey and classroom observation.

Note: Column A and Column B present group means that are regression-adjusted for the stratification used in the random assignment design. Column A-B presents differences in the regression-adjusted group means between the treatment and control groups. The table includes *p*-values from tests of differences between group means are also included in the table.

* Difference in group means is statistically significant at the .05 level.

B. Baseline equivalence of additional outcomes from principal survey

We found no statistically significant differences between treatment and control groups in principals' perceptions of safety in the community. We asked principals the same question we asked students about the frequency with which problems such as theft of belongings, verbal

fights, and physical fights occurred within the schools. Not surprisingly, reports on frequency from the students were higher than reports from principals. In third-cycle schools, only 2 percent of principals reported that thefts of belongings and physical fights occurred at least once a week. In addition, 21 percent of principals reported that verbal fights occurred at least once per week. Differences between study groups were not statistically significant. In secondary schools, less than 8 percent of principals reported that theft of belongings or physical fights occurred at least once a week and approximately 13 percent of principals reported that verbal fights occurred at least once a week. The difference between study groups is not statistically significant.

Table C.2. Principals reporting problems occurred at least once a week within schools (%)

	Treatment group (A)	Control group (B)	Difference (A-B)	*p*-value
Third-cycle				
Theft of belongings	2.0	3.0	-0.9	0.713
Physical fights between students	2.0	3.0	-0.9	0.713
Verbal fights between students	23.8	17.5	6.3	0.381
Number of principals	**54**	**62**		
Secondary				
Theft of belongings	3.0	7.0	-4.0	0.508
Physical fights between students	1.6	7.7	-6.2	0.107
Verbal fights between students	13.4	14	-0.7	0.943
Number of principals	**34**	**40**		

Source: 2017 principal survey.

Note: Column A and Column B present group means that are regression-adjusted for the stratification used in the random assignment design. Column A-B presents differences in the regression-adjusted group means between the treatment and control groups. The table includes *p*-values from tests of differences between group means are also included in the table.

* Difference in group means is statistically significant at the .05 level.

C. Baseline equivalence of additional outcomes from student survey

We found no statistically significant differences between study groups in the frequency with which problems such as theft of belongings and fights occurred in school. As shown in Table C.3, we found no statistically significant differences between study groups in the reported frequency of problems within the schools: 14 percent of students reported that theft of belongings occurred at least once a week, 22 percent of students reported that physical fights occurred at least once a week, and 30 percent reported that verbal fights occurred at least once a week. In secondary schools, approximately 16 percent of students reported that theft of belongings and physical fights occurred at least once a week. Finally, 40 percent of students in both types of schools reported that verbal fights occurred at least once a week. Differences between treatment and control groups are not statistically significant. Not surprisingly, the reports on frequency from the students were higher than the reports from principals. In third-cycle schools, only 2 percent of principals reported that thefts of belongings and physical fights occurred at least once a week. In addition, 21 percent of principals reported that verbal fights occurred at least once per week. Differences between study groups are not statistically

significant. In secondary schools, less than 8 percent of principals reported that theft of belongings or physical fights occurred at least once a week, approximately 13 percent of the principals reported that verbal fights occurred at least once a week. Differences between study groups are not statistically significant.

Table C.3. Students reporting problems occurred at least once a week within schools (%)

	Treatment group (A)	Control group (B)	Difference (A-B)	*p*-value
Third-cycle				
Theft of belongings	16.7	12.5	4.2	0.121
Physical fights between students	23.9	20.3	3.6	0.181
Verbal fights between students	30.9	28.5	2.4	0.46
Number of students	**539**	**417**		
Secondary				
Theft of belongings	17.3	15.6	1.6	0.642
Physical fights between students	17.7	17.4	0.3	0.923
Verbal fights between students	40.6	38.4	2.2	0.664
Number of students	**340**	**400**		

Source: 2017 student survey.

Note: Column A and Column B present group means that are regression-adjusted for the stratification used in the random assignment design. Column A-B presents differences in the regression-adjusted group means between the treatment and control groups. The table includes *p*-values from tests of differences between group means are also included in the table.

* Difference in group means is statistically significant at the .05 level.

We found no statistically significant differences in the percentage of students who expected to study a baccalaureate program (for third-cycle students) or complete it (for secondary students). We found differences in the percentage of students who expected to study a technical or professional career. More third-cycle students in the treatment group expected to study a professional career compared with the control group; this difference is statistically significant. We also found that more secondary students in the treatment group did not know if they would pursue a post-secondary education compared to the control group. The difference of 3 percentage points is statistically significant. Finally, we found no significant differences in the percentage of students interested in workshops; however, among secondary students, we found differences in the types of workshops of interest (Table C.4.)

Table C.4. Students expectations and interests (%)

	Treatment group (A)	Control group (B)	Difference (A-B)	*p*-value
Third-cycle schools				
Expect to study a baccalaureate program				
General	46.7	50.0	-3.3	0.432
Technical	47.8	42.5	5.3	0.194
Expect to study some technical or professional career				
Technical career	18.8	22.0	-3.2	0.18
Professional career	52.7	44.8	7.9*	0.034
Do not know	3.1	4.7	-1.6	0.177
Students interested in workshops	87.8	88.6	-0.8	0.73
Workshops of interest:				
English	18.8	19.1	-0.3	0.909
ITC	25.2	31.1	-5.9	0.112
Labor initiation	2.1	1.6	0.5	0.434
Vocational training	5.4	4.9	0.6	0.776
Soft skills	5.2	2.9	2.3	0.055
Art	18.4	20.8	-2.3	0.381
Sports and recreation	9.6	9.7	-0.1	0.953
Mechanics	8.3	9.6	-1.3	0.466
Other workshops	13.0	12.0	1.0	0.694
Number of students	**539**	**617**		
Secondary schools				
Expect to complete				
General	53.2	54.6	-1.4	0.879
Technical	46.8	45.4	1.4	0.879
Expect to study some technical or professional career	76.2	71.9	4.3	0.297
Technical degree	17.6	15.0	2.6	0.496
Professional degree	54.2	55.7	-1.5	0.727
Do not know	4.4	1.2	3.2*	0.014
Students interested in the workshops	90.5	90.3	0.2	0.937
Workshops of interest:				
English	22.7	27.1	-4.5	0.131
ITC	20.9	32.7	-11.8*	0.001
Labor initiation	4.1	3.2	0.9	0.608
Vocational training	8.5	3.9	4.6*	0.050
Soft skills	7.2	2.7	4.5*	0.011
Art	20.7	18.4	2.3	0.438
Sports and recreation	9.1	9.7	-0.6	0.808
Mechanics	4.1	8.2	-4.1*	0.029
Other workshops	17	12.6	4.4	0.237

TABLE C.4 *(Continued)*

	Treatment group (A)	Control group (B)	Difference (A-B)	*p*-value
Number of students	**340**	**400**		

Source: 2017 student survey.

Note: Column A and Column B present group means that are regression-adjusted for the stratification used in the random assignment design. Column A-B presents differences in the regression-adjusted group means between the treatment and control groups. The table includes *p*-values from tests of differences between group means are also included in the table.

ITC: Information Technology Center

* Difference in group means is statistically significant at the .05 level.

D. Baseline equivalence of additional outcomes from the teaching practices in the classrooms

As mentioned in Chapter III, we estimated the percentage of time teachers were absent from the room in each instance observed. In third-cycle school, half of the total time spent absent from the classroom occurred at the beginning of the class, when teachers' absence from the classroom usually occurred. In both groups, half of the time teachers were absent from the classroom occurred during the first two snapshots (4 percent of the time). In secondary schools, half of the total time teachers spent absent from the classroom occurred at the beginning of the class. We found a statistically significant difference between study groups in the share of time teachers were absent from the classroom during the first observation (4 percent in the treatment group compared to 2 percent in the control group).

Table C.5 Share of class time teachers were absent from the classrooms by observation

	Treatment group (A)	Control group (B)	Difference (A-B)	*p*-value
Share of time teachers were absent from the classroom in each snapshot or observation				
Third-cycle schools				
Absent 1st observation	2.9	2.6	0.3	0.563
Absent 2nd observation	1.3	1.0	0.2	0.503
Absent 3rd observation	0.6	0.8	-0.3	0.355
Absent 4th observation	0.5	0.5	-0.0	0.976
Absent 5th observation	0.4	0.6	-0.2	0.350
Absent 6th observation	0.1	0.3	-0.2	0.229
Absent 7th observation	0.3	0.5	-0.2	0.371
Absent 8th observation	0.6	0.5	0.1	0.556
Absent 9th observation	0.5	0.6	-0.1	0.722
Absent 10th observation	0.8	0.6	0.2	0.401
Secondary schools				
Absent 1st observation	3.8	2.4	1.4*	0.036
Absent 2nd observation	1.8	1.0	0.8	0.134
Absent 3rd observation	0.9	0.6	0.4	0.438

TABLE C.5 *(Continued)*

	Treatment group (A)	Control group (B)	Difference (A-B)	*p*-value
Absent 4th observation	0.5	0.7	-0.2	0.602
Absent 5th observation	0.8	0.2	0.5	0.214
Absent 6th observation	0.7	0.4	0.2	0.504
Absent 7th observation	0.3	0.7	-0.4	0.175
Absent 8th observation	0.4	0.8	-0.4	0.187
Absent 9th observation	0.6	0.5	0.1	0.728
Absent 10th observation	0.7	0.7	0.0	0.960

Source: 2017 classroom observations.

Note: Column A and Column B present group means that are regression-adjusted for the stratification used in the random assignment design. Column A-B presents differences in the regression-adjusted group means between the treatment and control groups. The table includes *p*-values from tests of differences between group means are also included in the table.

* Difference in group means is statistically significant at the .05 level.

E. Baseline equivalence of additional subgroup analysis

The tables below show the same student indicators presented for the full sample in Chapter IV. The first three columns show the baseline equivalence among male students (columns A, B, and A-B), the next three columns show the baseline equivalence among female students (columns C, D, and C-D), and the last column shows the baseline differences between males and females.

Students' use of time by gender

The results of the gender subgroup analysis were similar to what we found in the full sample for third-cycle and secondary schools. However, for females in third-cycle schools, we found one significant difference between groups: commuting time. Table C.6 summarizes the findings. We found no significant differences between groups on the time spent among activities for males and females, with one exception. Students in both groups reported spending between four and six hours in class or doing academic activities (four hours on average for third-cycle students and six hours for secondary students), followed by time spent on personal care[15] (two hours on average), watching television or socializing (two hours), and other activities such as recreation, homework, family tasks, and so on. The exception was commuting time for female students in third-cycle schools, where we found a statistically significant difference between females in the treatment and control groups (12 minutes). No statistically significant difference was found for males. Furthermore, the between-group difference for females is different from the between-group difference for males, and is statistically significant.

[15] Personal care includes activities like personal hygiene, eating, and preparing meals or snacks.

Table C.6. Third-cycle and secondary students' use of time, by gender

	Males			Females			Group differences between males and females (*p*-value)
	Treatment group (A)	Control group (B)	Difference (A-B)	Treatment group (C)	Control group (D)	Difference (C-D)	
Third-cycle schools							
Academic activities	3.9	3.5	0.3	3.6	3.6	-0.0	0.079
Personal care	2.1	2.2	-0.1	2.4	2.4	0.0	0.319
TV/computer	2.0	1.8	0.1	1.9	1.7	0.2	0.738
Socialization	1.6	1.6	-0.1	1.7	1.6	0.1	0.569
Homework	1.2	1.2	0.0	1.4	1.2	0.2	0.431
Commuting time	1.0	1.0	-0.0	1.0	0.8	0.2*	0.029*
Family chores	0.9	1.1	-0.2	1.6	1.7	-0.0	0.203
Unpaid work	0.7	0.8	-0.1	0.4	0.5	-0.1	0.684
Sports	0.5	0.5	0.0	0.1	0.2	-0.0	0.589
Paid work	0.4	0.2	0.2	0.1	0.1	0.0	0.142
At church	0.3	0.2	0.1	0.2	0.2	0.0	0.157
Other activity	0.1	0.1	-0.0	0.1	0.2	-0.1	0.434
Number of students	**268**	**309**		**271**	**308**		
Secondary schools							
Academic activities	6.1	5.8	0.3	6.3	5.5	0.8	0.098
Personal care	2.2	2.1	0.1	2.3	2.3	0.1	0.829
Socialization	1.7	1.9	-0.2	1.3	1.7	-0.4	0.625
Homework	1.5	1.5	-0.0	1.6	1.3	0.3	0.268
TV/computer	1.3	1.2	0.1	1.1	1.2	-0.0	0.397
Transportation	1.1	1.0	0.0	1.0	1.1	-0.1	0.265
Sports	0.4	0.4	-0.1	0.1	0.1	0.1	0.438
Family chores	0.4	0.4	-0.1	0.6	0.9	-0.3	0.199
Unpaid work	0.2	0.2	-0.0	0.2	0.3	-0.1	0.377
At church	0.2	0.2	0.0	0.2	0.3	-0.1	0.507
Other activity	0.1	0.1	0.0	0.1	0.1	-0.0	0.480
Paid work	0.0	0.1	-0.1	0.0	0.0	-0.0	0.677
Number of students	**171**	**195**		**169**	**205**		

Source: 2017 student survey.

Note: Columns A, B, C, and D present group means that are regression-adjusted for the stratification design with a regression. Columns A-B and C-D present differences in the regression-adjusted group means between the treatment and control groups for males and females, respectively. The *p*-values of the differences by gender are also included in the table.

* Difference in group means is statistically significant at the .05 level.

Except for time spent commuting, we found no significant differences between treatment and control groups for both males and females in the time students spent in

various places. As with the full sample, for the female and male subgroups, there are no significant differences between groups on the time spent at home, school, at other houses, or outside. However, third-cycle female students in the treatment group spent more time in transportation compared with female students in the control group. The difference of 0.2 percentage points between treatment and control females is statistically significant from the difference for males (Table C.7).

Table C.7. Third-cycle and secondary students' use of time one day, by gender (hours)

	Males			Female			Group differences between males and females (p-value)
	Treatment group (A)	Control group (B)	Difference (A-B)	Treatment group (C)	Control group (D)	Difference (C-D)	
Third-cycle schools							
At home	7.4	7.6	-0.2	8.6	8.4	0.2	0.228
At school	4.3	4.1	0.2	4.1	4.1	-0.0	0.286
Commuting time	1.0	1.0	-0.0	1.1	0.8	0.2*	*0.022
At other houses	0.5	0.4	0.1	0.4	0.5	-0.1	0.203
Outside	0.7	0.6	0.2	0.3	0.3	0.1	0.445
At work	0.5	0.6	-0.1	0.1	0.1	0.0	0.709
At a commercial place	0.1	0.0	0.0	0.1	0.1	-0.0	0.557
Another place	0.5	0.3	0.2	0.3	0.3	0.0	0.181
Number of students	**268**	**309**		**271**	**308**		
Secondary schools							
At home	6.1	6.1	0.0	6.2	6.7	-0.5	0.113
At school	7.1	6.7	0.4	7.1	6.3	0.9	0.104
In transportation	1.1	1.1	0.1	1.0	1.1	-0.1	0.259
At other houses	0.3	0.5	-0.1	0.2	0.4	-0.2*	0.630
Outside	0.5	0.6	-0.1	0.3	0.2	0.1	0.104
At work	0.0	0.0	0.0	0.0	0.0	-0.0	0.399
At a commercial place	0.0	0.1	-0.1	0.1	0.2	-0.1	0.898
Another place	0.3	0.2	0.0	0.3	0.3	-0.0	0.632
Number of students	**171**	**195**		**169**	**205**		

Source: 2017 student survey.

Note: Columns A, B, C, and D present group means that are adjusted for the stratification design with a regression. Columns A-B and C-D present differences in the regression-adjusted group means between the treatment and control groups for males and females, respectively. The *p*-values of the differences by gender are included in the table.

* Difference in group means is statistically significant at the .05 level.

Perception of teachers' support and engagement in school, by gender

We estimated differences in the perception of teaching support for females and males, and found no statistically significant differences between groups, with one exception.

Among male students in third-cycle schools, we found a significant difference in the teacher academic support subscale in favor of the control group. However, we found no differences in the magnitude of the differences for males compared to the differences for females. We also found no significant differences in behavioral and emotional engagement between treatment and control groups for males and females (Table C.8).

Table C.8. Perception of teachers' support and engagement in school one day, by gender (hours)

	Males			Female			Group differences between males and females (*p*-value)
	Treatment group (A)	Control group (B)	Difference (A-B)	Treatment group (C)	Control group (D)	Difference (C-D)	
Third-cycle schools							
Teacher academic support	4.6	4.7	-0.1*	4.6	4.7	-0.0	0.432
Behavioral engagement	4.0	4.0	0.0	4.2	4.1	0.1	0.183
Emotional engagement	4.5	4.5	0.0	4.5	4.5	0.0	0.927
Number of students	**268**	**309**		**271**	**308**		
Secondary schools							
Teacher academic support	4.5	4.6	-0.1	4.5	4.6	-0.0	0.694
Behavioral engagement	4.0	4.0	-0.0	4.1	4.2	-0.1	0.812
Emotional engagement	4.4	4.4	-0.0	4.5	4.5	-0.0	0.932
Number of students	**171**	**195**		**169**	**205**		

Source: 2017 student survey.

Note: Columns A, B, C, and D present group means that are regression adjusted for the stratification design with a regression. Columns A-B and C-D present differences in the regression-adjusted group means between the treatment and control groups for males and females, respectively. The *p*-values of the differences by gender are included in the table.

* Difference in group means is statistically significant at the .05 level.

Perception of classroom climate and teaching practices

We found no statistically significant differences in the percentage of students who agree on statements about classroom climate and gender equity between treatment and control groups for females and for males. In third-cycle and secondary schools, we found similar results in study groups for males and females on the percentage of students who agreed on statements about feeling comfortable sharing their experiences in class, disliking to participate in class discussions, and feeling marked for being male or female (Table C.9).

Table C.9. Students' perception of classroom climate and gender equity by gender (%)

	Males			Female			Group differences between males and females (*p*-value)
	Treatment group (A)	Control group (B)	Difference (A-B)	Treatment group (C)	Control group (D)	Difference (C-D)	
Third-cycle schools							
I feel comfortable sharing my own perspectives and experiences in class	95.0	92.3	2.8	96.8	93.8	3.0	0.943
I felt marked in class for being a man or a woman	24.8	29.8	-5.0	20.9	22.5	-1.6	0.521
I feel that I have to work harder than other students to be perceived as good	72.5	78.8	-6.3	67.4	72.5	-5.1	0.843
In class, I have heard teachers discriminate between men and women	8.3	12.5	-4.1	12.1	10.4	1.7	0.154
I do not like to participate in class discussions	36.9	39.2	-2.3	36.1	41.9	-5.7	0.491
Number of students	**268**	**309**		**271**	**308**		
Secondary schools							
I feel comfortable sharing my own perspectives and experiences in class	93.1	95.5	-2.4	93.0	95.3	-2.4	0.991
I felt marked in class for being a man or a woman	19.4	24.6	-5.1	20.3	15.2	5.2	0.067
I feel that I have to work harder than other students to be perceived as good	66.6	66.6	-0.0	61.9	51.5	10.4	0.135
In class, I have heard teachers discriminate between men and women	11.1	9.6	1.5	12.5	13.2	-0.7	0.656
I do not like to participate in class discussions	39.1	34.3	4.8	42.3	40.6	1.7	0.651
Number of students	**171**	**195**		**169**	**205**		

Source: 2017 student survey.

Note: Columns A, B, C, and D present group means that are regression adjusted for the stratification design with a regression. Columns A-B and C-D present differences in the regression-adjusted group means between the treatment and control groups for males and females, respectively. The *p*-values of the differences by gender are included in the table.

* Difference in group means is statistically significant at the .05 level.

Perception of community safety

For males and for females, we found no differences between groups on the perception of safety in the community, with one exception. In third-cycle schools, we found no statistically

significant differences between study groups in the perception of safety in the community for female and male students. In secondary schools, more male students in the control group reported feeling safe going to school compared to treatment males. This difference is statistically significant at the 5 percent level (Table C.10).

Table C.10. Percentage of students who reported feeling safe in the community

	Male			Female			Group differences between males and females (*p*-value)
	Treatment group (A)	Control group (B)	Difference (A-B)	Treatment group (C)	Control group (D)	Difference (C-D)	
Third-cycle schools							
Feel safe at school	88.0	88.7	-0.7	87.6	87.7	-0.1	0.858
Feel safe going to school	63.1	61.0	2.1	60.4	51.6	8.8	0.303
Feel safe in their community	78.2	75.7	2.5	71.0	72.8	-1.8	0.419
Number of students	**268**	**309**		**271**	**308**		
Secondary schools							
Feel safe at school	86.5	88.8	-2.4	86.0	91.7	-5.7	0.402
Feel safe going to school	42.7	60.6	-17.9*	46.8	49.6	-2.8	0.090
Feel safe in their community	65.1	75.4	-10.3	64.6	71.9	-7.3	0.711
Number of students	**171**	**195**		**169**	**205**		

Source: 2017 student survey.

Note: Columns A, B, C, and D present group means that are regression adjusted for the stratification design with a regression. Columns A-B and C-D present differences in the regression-adjusted group means between the treatment and control groups for males and females, respectively. The *p*-values of the differences by gender are included in the table.

* Difference in group means is statistically significant at the .05 level.

More females in the treatment group in secondary schools reported spending more than four hours working on mathematics, social science, and English tasks than females in the control group. We found no statistically significant differences between groups for males. Table C.11 shows that, in secondary schools, more female students in the treatment group reported spending more than four hours per week working on mathematics tasks compared to females in the control group. The difference of 10 percentage point is statistically significant. We also found a significant difference of 15 percentage points in favor of female students in the treatment group who spent more than four hours in social science tasks. Finally, we found that more female students in the treatment group spent more than four hours working on English tasks. These differences are statistically significant for females but not for males. Furthermore, the difference between treatment and control females is statistically significant from the difference for males.

Table C.11. Percentage of male and female students who reported spending more than four hours per week in subject area tasks

	Males			Females			Group differences between males and females (p-value)
	Treatment group (A)	Control group (B)	Difference (A-B)	Treatment group (C)	Control group (D)	Difference (C-D)	
Percentage of students who reported to spend more than 4 hours per week working on tasks in the following subject areas:							
Third-cycle schools							
Mathematics	16.6	21.1	-4.6	17.5	17.6	-0.2	0.272
Science, health, and environment	16.0	24.3	-8.2	18.6	23.7	-5.1	0.531
Social science or civics	15.7	23.2	-7.5*	16.0	21.2	-5.2	0.652
English	15.6	17.3	-1.7	17.0	16.8	0.2	0.657
Language or literature	15.0	18.8	-3.8	14.5	14.6	-0.1	0.465
Number of students	**268**	**309**		**271**	**308**		
Secondary schools							
Science, health, and environment	28.9	31.3	-2.4	30.1	22.6	7.6	0.187
Language or literature	24.0	24.4	-0.4	25.3	20.5	4.8	0.530
Mathematics	23.6	31.6	-8.0	35.0	24.8	10.2*	0.023*
Social science or civics	23.5	33.3	-9.8	33.4	18.5	14.9*	0.001*
English	18.9	22.7	-3.8	20.7	12.0	8.6*	0.023*
Number of students	**171**	**195**		**169**	**205**		

Source: 2017 student survey.

Note: Columns A, B, C, and D present group means that are regression adjusted for the stratification design with a regression. Columns A-B and C-D present differences in the regression-adjusted group means between the treatment and control groups for males and females, respectively. The *p*-values of the differences by gender are included in the table.

* Difference in group means is statistically significant at the .05 level.

For males and for females, we found no differences between groups in the percentage of students who reported that most of the teachers perform positive teaching practices, with one exception. In third-cycle schools, we found similar results in the majority of the teaching practices reported by male and female students in both study groups. Among males, more control students reported that most of their teachers share personal stories or experiences in class compared with treatment students; the difference of 8 percentage points is statistically significant. However, the differences for males compared to the differences for females are not statistically different. In secondary school, we also found similar results in the percentage of teaching practices reported by male and female students in both study groups (Table C.12).

Table C.12. Students' perception of teaching practices (%)

	Males			Female			Group differences between males and females (*p*-value)
	Treatment group (A)	Control group (B)	Difference (A-B)	Treatment group (C)	Control group (D)	Difference (C-D)	
Percentage of students who reported that most of their teachers:							
Third-cycle schools							
Encourage students with different characteristics to work together	96.2	95.2	1.0	94.5	94.4	0.1	0.749
Help students learn how to make positive changes in society	95.5	93.2	2.3	94.4	94.7	-0.4	0.438
Encourage students to contribute different points of view in class	93.6	91.5	2.2	92.8	95.9	-3.1	0.094
Motivate students to work harder than they thought they could	92.3	93.0	-0.7	92.6	94.2	-1.6	0.763
Teach students to be tolerant and respectful of different ideas or beliefs	90.4	92.7	-2.2	91.8	94.4	-2.6	0.901
Value the strengths and difficulties of the students in the classroom	90.4	88.9	1.5	88.4	92.1	-3.7	0.122
Take into account the different abilities of the students	89.8	86.9	2.9	89.1	89.0	0.0	0.446
Transform controversial issues into meaningful discussions for students	83.5	80.0	3.5	77.3	78.5	-1.2	0.311
Share in class some of their stories or personal experiences	73.4	81.3	-7.9*	77.9	79.8	-1.8	0.310
Speak openly about issues of social inequality	74.0	72.9	1.1	71.4	71.7	-0.3	0.825
Number of students	**268**	**309**		**271**	**308**		
Secondary schools							
Encourage students with different characteristics to work together	94.2	90.5	3.7	91.5	93.0	-1.5	0.170
Encourage students to contribute different points of view in class	92.0	90.5	1.5	93.4	90.2	3.2	0.680
Teach students to be tolerant and respectful of different ideas or beliefs	89.6	92.7	-3.1	91.4	93.9	-2.5	0.901
Motivate students to work harder than they thought they could	88.5	93.7	-5.2	92.2	90.8	1.4	0.101
Take into account the different abilities of the students	87.3	89.4	-2.1	84.0	85.8	-1.7	0.934
Help students learn how to make positive changes in society	86.1	91.6	-5.5	87.9	91.3	-3.4	0.600
Value the strengths and difficulties of the students in the classroom	85.7	90.4	-4.7	77.8	87.0	-9.2	0.370
Transform controversial issues into meaningful discussions for students	78.1	78.5	-0.4	68.5	69.3	-0.8	0.946
Speak openly about issues of social inequality	73.0	79.1	-6.0	77.5	75.4	2.2	0.172

TABLE C.12 *(Continued)*

	Males			Female			Group differences between males and females (*p*-value)
	Treatment group (A)	Control group (B)	Difference (A-B)	Treatment group (C)	Control group (D)	Difference (C-D)	
Number of students	**171**	**195**		**169**	**205**		

Source: 2017 student survey.

Note: Columns A, B, C, and D present group means that are regression adjusted for the stratification design with a regression. Columns A-B and C-D present differences in the regression-adjusted group means between the treatment and control groups for males and females, respectively. The *p*-value of the differences by gender are included in the table.

* Difference in group means is statistically significant at the .05 level.

The results of the gender subgroup analysis are similar to what we found in the full sample. As shown in Table C.13, we found no statistically significant differences in the frequency of problems such as theft of belongings and fights occurring in schools between study groups by gender.

Table C.13. Students who reported problems occurred at least once a week within schools

	Males			Female			Group differences between males and females (*p*-value)
	Treatment group (A)	Control group (B)	Difference (A-B)	Treatment group (C)	Control group (D)	Difference (C-D)	
Percentage of students who reported that the following problems occurred at least once a week							
Third-cycle schools							
Theft of belongings	16.9	13.4	3.6	16.5	11.7	4.8	0.762
Physical fights between students	19.8	18.8	1.0	28.0	21.9	6.1	0.343
Verbal fights between students	28.6	27.4	1.2	33.1	29.5	3.6	0.641
Number of students	**268**	**309**		**271**	**308**		
Secondary schools							
Theft of belongings	14.6	13.6	0.9	20.0	17.5	2.6	0.773
Physical fights between students	13.3	12.5	0.8	22.1	21.9	0.3	0.931
Verbal fights between students	40.2	42.4	-2.2	41.0	34.8	6.2	0.205
Number of students	**171**	**195**		**169**	**205**		

Source: 2017 student survey.

Note: Columns A, B, C, and D present group means that are adjusted for the stratification design with a regression. Columns A-B and C-D present differences in the regression-adjusted group means between the treatment and control groups for males and females, respectively. The *p*-values of the differences by gender are included in the table.

* Difference in group means is statistically significant at the .05 level.

The results of the gender subgroup analysis are similar to what we found in the full sample. For males and females, we found no statistically significant differences between study groups in the percentage of students who expected to study a baccalaureate program (for third-cycle students) or complete it (for secondary students). We found no significant differences between treatment and control in the percentage of students who expected to study a technical or professional career for each gender. Finally, we found significant differences in the interest reported in ITC workshops for secondary students. Although we found that the differences between treatment group and control group males are greater than the differences between treatment group and control group females, the differences are not statistically significant (Table C.13.)

Table C.13. Students expectations and interests by gender

	Males			Females			Group differences between males and females (*p*-value)
	Treatment group (A)	Control group (B)	Difference (A-B)	Treatment group (C)	Control group (D)	Difference (C-D)	
Third-cycle schools							
Expect to study a baccalaureate (%)							
General	48.1	50.2	-2.1	45.3	49.7	-4.4	0.692
Technical	46.7	43.6	3.1	48.9	41.4	7.5	0.441
Expect to study some technical or professional career (%)							
Technical degree	20.1	22.7	-2.6	17.5	21.3	-3.8	0.803
Professional degree	51.6	41.2	10.4	53.7	48.4	5.3	0.511
Do not know	3.8	5.5	-1.7	2.4	3.8	-1.4	0.898
Student interested in workshops (%)	87.0	90.0	-3.0	88.7	87.2	1.4	0.376
Workshop of interest:							
English	16.2	18.0	-1.7	21.3	20.2	1.1	0.623
ITC	26.6	32.7	-6.1	23.8	29.4	-5.6	0.933
Labor initiation	2.0	2.0	0.1	2.1	1.1	1.0	0.521
Vocational training	5.5	2.1	3.5	5.3	7.7	-2.3	0.145
Soft skills	5.0	2.4	2.6	5.4	3.3	2.1	0.815
Art	12.1	16.6	-4.6	24.7	24.9	-0.2	0.339
Sports and recreation	10.6	10.8	-0.2	8.6	8.6	0.0	0.942
Mechanics	14.1	18.3	-4.2	2.6	0.8	1.8	0.094
Other workshops	9.2	7.2	2.0	16.7	16.8	-0.2	0.650
Number of students	**268**	**309**		**271**	**308**		
Secondary schools							
Expect to complete the baccalaureate (%)							
General	53.7	53.9	-0.2	52.7	55.3	-2.6	0.464
Technical	46.3	46.1	0.2	47.3	44.7	2.6	0.464
Expect to study some technical or professional career (%)							
Technical degree	21.6	16.3	5.3	13.5	13.8	-0.3	0.327
Professional degree	53.7	54.1	-0.5	54.8	57.2	-2.4	0.843
Do not know	4.6	1.2	3.4	4.2	1.2	3.0	0.877

TABLE C.13 *(Continued)*

	Males			Females			Group differences between males and females (*p*-value)
	Treatment group (A)	Control group (B)	Difference (A-B)	Treatment group (C)	Control group (D)	Difference (C-D)	
Student interested in workshops	87.3	90.7	-3.4	93.8	90.0	3.8	0.129
Workshops of interest:							
English	22.8	25.1	-2.3	22.5	28.9	-6.4	0.542
ITC	21.1	33.2	-12.1*	20.6	32.3	-11.6*	0.947
Labor initiation	4.0	2.7	1.3	4.1	3.7	0.4	0.720
Vocational training	8.2	3.8	4.4	8.8	4.0	4.8	0.920
Soft skills	7.2	2.1	5.1	7.3	3.3	4.0	0.766
Art	19.4	15.6	3.8	22.1	21.1	1.0	0.626
Sports and recreation	11.4	10.8	0.5	6.8	8.6	-1.8	0.540
Mechanics	8.6	15.2	-6.6	-0.6	1.8	-2.4	0.327
Other workshops	13.1	8.6	4.6	21.0	16.3	4.6	0.991

Source: 2017 student survey.

Note: Columns A, B, C, and D present group means that are regression adjusted for the stratification design with a regression. Columns A-B and C-D present differences in the regression-adjusted group means between the treatment and control groups for males and females, respectively. The *p*-values of the differences by gender are included in the table.

* Difference in group means is statistically significant at the .05 level.

APPENDIX D: EDUCATIONAL OUTCOMES FOR THIRD-CYCLE AND SECONDARY STUDENTS ENROLLED IN SCHOOL YEAR 2016 (BASE YEAR)

This page has been left blank for double-sided copying.

In this appendix, we present the educational outcomes for school years 2016 and 2017. These outcomes show similar results compared to the outcomes using student data for school years 2017 and 2018 previously described in Chapter V.

A. Baseline equivalence in third-cycle educational outcomes in base year 2016

Table D.1 shows the number of students enrolled in third-cycle schools in our study sample in the 2016 school year with data from school years 2016 and 2017.

Table D.1. Number of students enrolled by study group at the beginning of the 2016 school year

	Treatment group	Control group
Grade 7	6,499	7,159
Grade 8	6,006	6,662
Grade 9	5,604	6,281
Total number of students	18,109	20,102
Number of secondary schools	171	204
Number of systems	45	54

Source: MINED enrollment registry for 2016 (SAE system).

In third-cycle schools, the study groups had similar percentages of overage students. In both groups, the percentage of students who were overage in grades 7, 8, and 9 was similar. Approximately 15 percent of the students were overage in grade 7, 14 percent in grade 8, and 13 percent in grade 9.

In third-cycle students, the study groups had similar percentages of female students. The percentage of female students in the study groups is approximately 47 percent in grades 7, 8, and 9. The differences between groups are not statistically significant.

Table D.2. Students' characteristics in base year 2016

	Treatment group (A)	Control group (B)	Difference (A-B)	*p*-value
Percentage of overage students				
Grade 7	15.0	15.4	0.4	0.70
Grade 8	12.7	14.3	1.6	0.13
Grade 9	12.2	12.9	0.7	0.44
Percentage of female students				
Grade 7	46.6	44.8	-1.8	0.15
Grade 8	47.0	46.8	-0.1	0.91
Grade 9	47.3	48.1	0.8	0.49

Source: MINED (SAE system) for school year 2016

Note: Columns A and B present group means that are regression adjusted for the stratification used in the random assignment design. Column A-B presents differences in the regression-adjusted group means between the treatment and control groups. The *p*-values from tests of differences between group means are also included in the table.

* Difference in group means is statistically significant at the .05 level.

In third-cycle schools, educational outcomes such as progression to the next grade, dropout across years, and grade repetition were similar in both groups. In Table D.3, we present outcomes of the children enrolled in third-cycle schools at the beginning of the 2016 school year by the next school year, 2017. Out of students enrolled in grade 7 at the beginning of 2017, 73 progressed to grade 8 in the same school in the next school year, 22 percent dropped out of the same school in the next school year, and 6 percent repeated grade 7 in the same school in the next school year. Note that to be identified as having progressed to the next grade, a student must have passed the grade and enrolled in the next grade in the same school in the next year. Of the students enrolled in grade 8 in the 2016 school year, 76 progressed to grade 9 in the same school in 2017, 21 percent dropped out from the school, and 4 percent repeated grade 8 in the same school in the next school year. We found no statistically significant differences between treatment and control groups.

Transition from third-cycle to secondary was similar in both groups. The transition between third-cycle and secondary occurs from 9th grade to 10th grade. We found no statistically significant differences between groups in the percentage of students enrolled in grade 9 in 2016 who were enrolled in grade 10 in any secondary school in school year 2017.

Table D.3. Education outcomes from 2016 school year to 2017, by study group

	Treatment group (A)	Control group (B)	Difference (A-B)	*p*-value
Dropout rate from the school (across years)				
Grade 7	21.5	22.0	0.5	0.698
Grade 8	20.6	20.4	-0.2	0.881
Progressed to the next grade (in the same school)				
Grade 7	73.3	72.0	-1.3	0.362
Grade 8	75.8	75.7	-0.1	0.964
Repeated grade (in the same school)				
Grade 7	5.2	6.0	0.8	0.282
Grade 8	3.6	3.9	0.3	0.644
Grade 9	1.6	1.6	0.0	0.958
Transition from 9th to 10th grade				
Percentage of students enrolled in 2017 in grade 9 who were enrolled in grade 10 in any school in school year 2018	73.6	69.8	-3.8	0.148

Source: MINED (SAE system) for school year 2016

Note: Column A and Column B present group means that are regression-adjusted for the stratification used in the random assignment design. Column A-B presents differences in the regression-adjusted group means between the treatment and control groups. The table includes *p*-values from tests of differences between group means are also included in the table.

* Difference in group means is statistically significant at the .05 level.

B. Baseline equivalence in secondary educational outcomes in base year 2016

Table D.4 presents the enrollment disaggregated by type of baccalaureate for the schools in our sample for which MINED provided us data for the 2017 school year.

Table D.4. Number of students enrolled at the beginning of school year 2017, by study group

	Treatment group	Control group
General		
Grade 10	1,991	2,221
Grade 11	1,541	1,850
Technical		
Grade 10	1,507	2,611
Grade 11	1,377	1,890
Grade 12	1,198	1,855
Total number of students	7,614	10,434
Number of secondary schools	39	51
Number of systems	33	40

Source: MINED enrollment registry for 2016 (SIRAI system).

In secondary schools, the likelihood of being overage was similar in both groups and study groups had similar percentages of females, with one exception. In grade 10, approximately 12 percent of students in general programs and 11 percent in technical programs were overage. In grade 11, 12 percent of students in general programs and 10 percent in technical programs were overage. The differences between groups are not statistically significant. In general programs, the percentage of female students in the study groups was approximately 47 percent in grade 10. The difference between groups is not statistically significant. However, we found a significant difference in the percentage of female students in 11th grade: 51 percent in the treatment group compared to 47 percent in the control. In technical programs, study groups had similar percentages of females in grade 10 (48 percent), and grades 11 and 12 (50 percent each). The differences between groups are not statistically significant (Table D.5).

Table D.5. Students' characteristics by study group in base year (2016)

	Treatment group (A)	Control group (B)	Difference (A-B)	*p*-value
Percentage of overage students				
General				
Grade 10	12.2	12.3	0.1	0.908
Grade 11	11.0	13.5	2.5	0.099
Technical				
Grade 10	10.4	11.0	0.6	0.716
Grade 11	9.4	11.1	1.7	0.214

TABLE D.5 *(Continued)*

	Treatment group (A)	Control group (B)	Difference (A-B)	*p*-value
Grade 12	8.9	9.8	0.9	0.425
Percentage of female students				
General				
Grade 10	47.1	46.5	-0.6	0.730
Grade 11	51.3	46.7	-4.5*	0.040
Technical				
Grade 10	49.4	47.4	-2.0	0.155
Grade 11	48.5	50.9	2.4	0.207
Grade 12	49.7	50.7	1.0	0.683

Source: MINED (SIRAI system) for school year 2016.

Note: Column A and Column B present group means that are regression-adjusted for the stratification used in the random assignment design. Column A-B presents differences in the regression-adjusted group means between the treatment and control groups. The table includes *p*-values from tests of differences between group means are also included in the table.

* Difference in group means is statistically significant at the .05 level.

In secondary grades, progression to the next grade in the same school was similar across groups. A student who passes a grade can enroll in the next grade in the next school year, but not all students who pass enroll in the same school in the next year. We identified a student as having progressed to the next grade if the student enrolled in the next grade in the same school in the next school year. As shown in Table D.6, among students enrolled in 2016 in grade 10, approximately 73 percent progressed to grade 11 in the same school in the 2017 school year and 85 percent of students enrolled in grade 11 in technical programs progressed to grade 12 in 2017. The differences between treatment and control groups are not statistically significant.

Dropout rates were similar across groups. Among students enrolled in 2016 in grade 10 in general program, 24 percent of the students dropped out from that school in the next school year, 2017. The differences between control and treatment groups is not statistically significant. We also found no significant differences among study groups in dropout rates across years among students enrolled in grades 10 and 11 in technical programs.

Grade repetition rates were similar in both groups. In grade 10, repetition rate was 5 percent in general programs and technical programs in both groups. In grade 11, repetition rate was 2 percent among students in both groups in general and technical programs. Any differences between study groups are not statistically significant at the 5 percent level.

We found no statistically significant differences between groups in the percentage of students who passed the grade at the end of the school year in secondary schools. Approximately 78 percent of students enrolled in grade 10 in 2016 passed the grade in both types of program. Note that this outcome only assesses if a student passed the grade, regardless of whether the student enrolled in the next grade in the next school year. Among students enrolled in grade 11, 87 percent passed the grade in general and technical programs. Almost all students who were enrolled in grade 12 in technical programs passed that grade (Table D.6).

Table D.6. Education outcomes by study group in base year 2016

	Treatment group (A)	Control group (B)	Difference (A-B)	*p*-value
Dropout rate (across years) in the same school at the end of the school year				
General				
Grade 10	24.0	23.1	-1.0	0.599
Technical				
Grade 10	21.4	20.7	-0.7	0.806
Grade 11	14.6	12.1	-2.5	0.227
Progression to the next grade in the same school at the end of the school year				
General				
Grade 10	69.3	73.4	4.1	0.100
Technical				
Grade 10	73.9	73.7	-0.2	0.954
Grade 11	83.4	85.6	2.1	0.326
Repeated grade (in the same school)				
General				
Grade 10	6.7	3.6	3.1	0.147
Grade 11	2.3	1.8	0.5	0.463
Technical				
Grade 10	4.7	5.6	-0.9	0.550
Grade 11	2.0	2.4	-0.4	0.514
Grade 12	0.7	0.4	0.4	0.248
Passed rate (at the end of the school year)				
General				
Grade 10	75.2	80.0	-4.9	0.054
Grade 11	87.6	85.5	2.1	0.302
Technical				
Grade 10	78.3	78.5	-0.2	0.949
Grade 11	87.6	87.0	0.6	0.760
Grade 12	98.1	96.9	1.2	0.210

Source: MINED (SIRAI system) for beginning and end of school year 2016.

Note: Column A and Column B present group means that are regression-adjusted for the stratification used in the random assignment design. Column A-B presents differences in the regression-adjusted group means between the treatment and control groups. The table includes p-values from tests of differences between group means are also included in the table.

* Difference in group means is statistically significant at the .05 level.

We found no significant differences between study groups in the scores of the PAES standardized test in 2016 in any of the subject areas assessed. In both groups, the global score was approximately 5; the higher score was obtained in social science (5.8), followed by language and science (5.4), and mathematics (4.6). The differences between study groups are not statistically significant.

Table D.7. PAES score by study group in 2016

	Treatment group (A)	Control group (B)	Difference (A-B)	*p*-value
Global	5.0	5.1	-0.1	0.760
Mathematics	4.6	4.9	-0.3	0.101
Social science	5.6	5.5	0.1	0.482
Science	5.4	5.3	0.0	0.925
Language	5.4	5.4	0.1	0.484
Number of students	**2,571**	**3,266**		

Source: MINED (SIRAI system) for school year 2016

Note: Column A and Column B present group means that are regression-adjusted for the stratification used in the random assignment design. Column A-B presents differences in the regression-adjusted group means between the treatment and control groups. The table includes p-values from tests of differences between group means are also included in the table.

* Difference in group means is statistically significant at the .05 level.

This page has been left blank for double-sided copying.

www.mathematica-mpr.com

Improving public well-being by conducting high quality, objective research and data collection

PRINCETON, NJ ■ ANN ARBOR, MI ■ CAMBRIDGE, MA ■ CHICAGO, IL ■ OAKLAND, CA ■ SEATTLE, WA ■ TUCSON, AZ ■ WASHINGTON, DC ■ WOODLAWN, MD